

En undersøkelse hvorvidt systemer  
av stokastiske differensialligninger  
egner seg for å undersøke utvikling  
av, og interaksjon mellom,  
inntektsulikhet og  
velferdsstatsgenerøsitet

*Et empirisk studium*

Aslak Bakke Kvinlog



Master of philosophy and economics

UNIVERSITETET I OSLO

[18.10.2010]

# 1 Forord

To spesifikke hendelser ligger til grunn for denne oppgaven. Kalle Moene spurte Tore Schweder om et forslag til en stokastisk modellering for en problemstilling han da arbeidet på sammen med Erling Barth. Uavhengig av dette gikk jeg til Tore Schweder for å spørre om noen statistiske metoder. Jeg hadde hatt han som foreleser i *ECON5160 - Stochastic modeling and analysis*, og kjente ham derfra. Samtalen utviklet seg til at han kunne være min veileder, og at jeg kunne delta i prosjektet å finne en stokastisk modell. Prosjektet Kalle Moene og Erling Barth jobbet med handlet om interaksjon mellom inntektsulikheten og graden av velferdsstatsgenerøsitet. Resultatet er å lese i artikkelen *The Equality Multiplier*[Barth, Moene, 2009], og jeg vil først gi et sammendrag av denne. Deretter skisserer jeg den grunnleggende tankegangen som ligger bak den alternative modelleringen. Det matematiske grunnlaget for denne er stokastiske differensialligninger, eller mer spesifikt Ornstein-Uhlenbeck-prosesser. De umiddelbare statistiske egenskapene blir utregnet, og antagelser blir begrunnet. Mer perifere forutsetninger er lagt til et appendix. Avslutningsvis setter jeg opp spesifikke modeller og forklarer hvordan den numeriske analysen blir gjennomført. Grunnlaget for denne er sannsynlighetsmaksimeringsteorien, og programmeringsarbeidet er gjort i C.

Jeg vil takke min veileder, Tore Schweder, for hjelp og velvillighet hver gang jeg kom på kontoret. Det var stort sett plutselig og uten avtale, og jeg var alltid velkommen for diskusjon og forklaringer. Det var en forutsetning for å kunne gå videre. Jeg har vært helt avhengig av min biveileder, Trond Reitan, til denne oppgaven. Han har samarbeidet med Tore Schweder tidligere, og kjente godt til det teoretiske grunnlaget for arbeidet. Oppgaven ville stoppet med bare analytiske uttrykk og spekulasjoner om det ikke hadde vært for han. Jeg har fått langt mer hjelp av han til det numeriske arbeidet, men også det teoretiske, enn hva jeg kunne forvente. Hans generøsitet har vært til stor motivasjon, og jeg er grunnleggende takknemlig. Jeg har også fått hjelp av Atle Seierstad og Nils Christian Framstad, og jeg har satt stor pris på det.

Til oppgaven har jeg fått ESOP-stipendiet og jeg takker for dette.

# Innhold

<b>1</b>	<b>Forord</b>	<b>1</b>
<b>2</b>	<b>Sammendrag</b>	<b>4</b>
<b>3</b>	<b>Introduksjon</b>	<b>2</b>
3.1	<i>The Equality Multiplier</i> . . . . .	2
3.2	Grunnlaget for alternativ modellering . . . . .	4
3.3	Ligningssystemer . . . . .	7
3.4	Sannsynlighetsmaksimering . . . . .	10
3.5	Likelihood-funksjonen . . . . .	11
<b>4</b>	<b>Den grunnleggende tankegangen i modelleringen</b>	<b>21</b>
<b>5</b>	<b>Beskrivelse av den numeriske analysen</b>	<b>23</b>
<b>6</b>	<b>Modellering</b>	<b>25</b>
6.1	Referansemodell . . . . .	25
6.1.1	Modell 1 . . . . .	28
6.1.2	Modell 2 . . . . .	29
6.2	Modell 3 . . . . .	32
<b>7</b>	<b>Konklusjon</b>	<b>35</b>
<b>8</b>	<b>Appendix</b>	<b>36</b>
8.1	Førsteordens lineære differensialligninger . . . . .	36
8.1.1	Deterministisk . . . . .	36
8.1.2	Stokastikk . . . . .	37
8.2	Momenter for en én-dimensjonal Ornstein-Uhlenbeck-prosess .	38
8.3	Standardisert Wiener-prosess . . . . .	40
8.4	The Ito formula . . . . .	43
8.5	Maximum likelihood . . . . .	43
8.6	Optimeringsutskrifter for modell 2 . . . . .	44
8.7	Resultater for ett og to lag, med og uten interaksjon . . . . .	45
8.8	C-koden . . . . .	48



## 2 Sammendrag

Oppgaven er et forsøk på remodellering av problemstilling Kalle Moene og Erling Barth stiller i artikkelen *The Equality Multiplier* (Barth, Moene 2009), hvor interaksjonen mellom inntektsulikhet og velferdsstatsgenerøsitet studeres. Denne interaksjonen modelleres ved hjelp av systemer av Ornstein-Uhlenbeck-prosesser, som er stokastiske differensialligninger. Jeg ser for meg at prosessene utvikler seg i kontinuerlig tid, og at de gjensidig påvirker hverandre. Observasjonene er hentet fra spesifikke år, men dette betyr bare at observasjonene skjer på spesifikke tidspunkt, og at dette ikke har noe med prosessen i seg selv å gjøre. Datamaterialet blir remodellert for å passe inn i den sannsynlighetsmodellen som modellen forutsetter, og analysen blir gjort i et C-program som er spesialskrevet av Trond Reitan, og lettere modifisert av meg. Datamaterialet jeg bruker er tall for inntektsulikhet, målt som niende på første desil i timelønnsfordelingen, og velferdsstatsgenerøsitet, målt etter en indeks[Scruggs] som er utarbeidet av Lyle Scruggs, University of Connecticut, alt fra perioden 1976-2002.

Endogenitetsproblemer kan i lineær analyse korrigeres for ved hjelp av instrumentvariable. I denne oppgaven blir endogeniteten *endogenisert*, slik at interaksjonsparametrene opptre som ordinære parametre i likelihood-funksjonen, uten at instrumentvariable trekkes inn. Etter passende transformasjoner av datamaterialet, gjør jeg en antakelse om at disse er bestemt av de stokastiske differensialligningene, slik at data er normalfordelt. Et histogram viser at antagelsen kan være brukbar.

Systemets kovariansmatrise blir utregnet og danner grunnlaget for optimeringen, som gjøres etter sannsynlighetsmaksimeringsprinsippet. Optimeringsrutinen søker det parametersettet som best forklarer variasjonen i det observerte datamaterialet. Når disse parametrene blir sett i lys av den grunnleggende modellen igjen, så gir de grunnlag for kvantitativ og kvalitativ analyse.

Dynamiske modeller bidrar til å besvare spørsmål som statiske modeller ikke gjør, nemlig spørsmål om tiden. Jeg gjør en analyse over utviklingen i inntektsulikhet og velferdsstatsgenerøsitet, og ser på hvordan de estimerte

parametrene kan si noe om fleksibiliteten eller tregheten i disse størrelsene, for forskjellige politiske systemer.

Resultatene er både positive og negative, altså i seg selv. Det lot seg ikke gjøre å estimere de optimale interaksjonsparametrene, men resultatene var likevel tydelige på at modeller med interaksjon *bedre* forklarte datamaterialet, enn modeller *uten*. Likelihooden var for alle modeller preget av vesentlig multimodalitet, altså flertoppethet. Problemet var større for kompliserte modeller enn for enkle, slik en kan forvente, og multimodaliteten ble ekstrem i modeller med interaksjon. Siden prosessene er modellert ved stokastiske differensialligninger, går det an å regne ut kovariansen mellom observasjoner fra prosessen på to forskjellige tidspunkt. Denne er avhengig av modellens parametre, og komplette utregninger av denne kovariansen er gjort i oppgaven. Forskjellige parametre gir forskjellig kovarians, og noen parametre vil gi en teoretisk kovarians som passer bedre med den observerte kovariansen, enn andre. I optimeringen søker jeg spesifikt de parameterverdiene som har den egenskapen at den teoretiske kovariansen *best mulig* forklarer den observerte kovariansen. Frekventistisk inferens ble brukt, men i etterpåklokskapens lys kan det tenkes at Bayesiansk MCMC-optimering ville ført frem. Det får stå til senere å finne ut av. I frekventistisk analyse gjøres det ingen streng antakelse om parametrene i forkant. Et sett parameterverdier blir trukket, basert på en hensiktsmessig transformering av en standard uniform fordelt trekning. Basert på dette sett av parameterverdier regnes likelihooden ut, og optimeringen går ut på å *finne veien* gjennom parameterrommet, frem til det parametersettet som maksimerer likelihooden, og veien dertil er gradvis, altså parametrene endres gradvis. Hvis det fra et gitt utgangspunkt ikke er mulig å finne parametre i "*nabolaget*" som bedre forklarer observasjonene, så konkluderes det med at dette er det beste parametersettet. Hvis likelihooden er flertoppet betyr det at optimeringen gjerne konkluderer med et parametersett som er *lokalt best*, men gjerne ikke *globalt best*. Dette blir forsøkt *korrigert* ved å kjøre mange separate optimeringer, hver med sin individuelle trekning av parametre som utgangspunkt. Tanken er at en ved mange kjøringar til slutt finner veien til toppen. En betydelig ulempe med metoden er at den ikke gir noe estimat på usikkerheten i selve esti-

matet. Kurvaturen til log-likelihooden i topp-punktet gir dårlig informasjon om dette. En kunne brukt parametrisk bootstrappingmetoder, men det er ressurskrevende og ikke gjort.

Modeller med forskjellig grad av kompleksitet blir forsøkt, både *med* den omtalte interaksjonen, og *uten*. I modeller uten interaksjon reduseres modellen til to separate modeller; én for inntektsulikhet, og én for velferdsstatsgenerøsitet. Dette tillater interessante studier i seg selv, og kan besvare spørsmål om tidsaspekter i trenden av hver av disse. Likelihooden i disse modellene er bare i liten grad preget av flertoppethet, og flertallet av optimeringene kommer til samme konklusjon for parametrene. I modeller med interaksjon blir kovariansstrukturen veldig mye mer komplisert; en stor sum av ikke-lineære kombinasjoner av parametrene. En liten endring i parametrene vil da kunne føre til stor endring i kovariansverdien, og dette er trolig årsaken til at modeller med interaksjon er preget av vesentlig flertoppethet. Det vil altså si at optimeringen i et hvert tilfelle fant et parametersett som *best i sitt "nabolag"* forklarte observasjonene. Men dette nabolaget var stort sett i umiddelbar nærhet til den initiale trekningen, iallefall for flere av parametrene, og gir ikke grunnlag for økonomisk tolkning. Altså økonomisk tolkning av hva som ville vært fornuftige parameterverdier gir grunnlag for å hevde dette.

I modellene uten interaksjon var det altså mulig å komme til estimerer på parametrene, og dermed til økonomiske konklusjoner. Jeg baserer meg via *The Equality Multiplier* [Barth, Moene, 2009] indirekte på *The Three Worlds of Welfare Capitalism* [Esbing-Andersen, 1990], som inndeler land i tre kategorier: Konservative land, sosialdemokratiske land og liberale land. Jeg finner at *stabiliteten i utviklingen* av inntektsulikhet og velferdsstatsgenerøsitet er mindre for de konservative landene. Det vil si at veksten i disse størrelsene kan ligge flere år under eller over trend, før den henter seg inn igjen. Både de liberale landene og de sosialdemokratiske landene bruker kort tid på å innhente trenden.

## 3 Introduksjon

### 3.1 *The Equality Multiplier*

Kalle Moene og Erling Barth viser i artikkelen *The Equality Multiplier* [Barth, Moene, 2009] hvordan inntekstulikhet og velferdsstatsgenerøsitet gjensidig påvirker hverandre. De ser på to separate effekter, hvor den ene er inntekstulikhetens påvirkning på velferdsstatsgenerøsitet, og den andre er den motsatte. Begrunnelsen for at velferdsstatsgenerøsitet påvirker inntekstulikheten er at det er de svakeste gruppene i samfunnet som trenger støtten mest. Velferdsstaten fungerer som et sikkerhetsnett som reduserer risikoen ved arbeidsledighet, og gjør det dermed lettere å kreve høyere lønn. Denne mekanismen gjør at inntekstulikheten blir redusert fra nedsiden, altså at lønnen under medianen blir mer sammentrykket. Lønnsfordelingen i det høyere inntekstsskiktet er mer spredt, og totalresultatet er at medianlønnen ligger under gjennomsnittslønnen. Velferdsstaten ble i utgangspunktet utviklet som en sosial forsikring for enkeltindividet, og denne lønnsutjevningstendensen er en sekundær effekt i så måte. Begrunnelsen for den andre effekten, som går fra inntekstulikhet til velferdsstatsgenerøsitet, er sosial identifikasjon og egne sikkerhetsbehov. Ved lav inntekstulikhet, hvor sammenpressing skjer fra undersiden, vil nettopp flertallet av velgerne befinne seg under gjennomsnittet i lønnsfordelingen. Dette flertallet vil dermed på rasjonelt grunnlag velge å støtte statens velferdsordninger.

Selve modelleringen er gjort på følgende måte:

$$\begin{aligned} \ln(G_j) &= A_j - a_I \ln(I_j), \quad A_j = A(z_j) \\ \ln(I_j) &= B_j - a_G \ln(G_j), \quad B_j = B(y_j) \end{aligned}$$

hvor  $G_j$  er velferdsstatsgenerøsiteten, og  $I_j$  er inntekstulikheten, som er definert som niende desil på første desil,  $\frac{d_9}{d_1}$ , i timelønn før skatt, begge for land  $j$ . Interessen her er knyttet til parametrene  $a_I < 0$  og  $a_G < 0$ . Siden størrelsene er log-transformerte, blir interesseparametrene å tolke som elastisiteter<sup>1</sup>:  $a_I$  er altså elastisiteten i  $G_j$  med hensyn på  $I_j$ , og omvendt for  $a_G$ .  $z_j$  er en

---

<sup>1</sup> $El_y x = \frac{y}{x} \frac{dx}{dy} = \frac{d(\ln(x))}{d(\ln(y))} \Rightarrow d(\ln(x)) = El_y(x) d(\ln(y))$



vektor som inkluderer politisk farge på det styrende partiet, landets inntektsnivå og økonomiske risikofaktorer for velgerne, som f. eks. grad av åpenhet i økonomien.  $y_j$  inkluderer indikatorer for lønnsfastsettelsessystemet, fagforeningstetthet og landets inntektsnivå<sup>2</sup>. Dermed er  $A_j$  og  $B_j$  konstanter med spesifikke verdier knyttet til tidspunktet.

En endring i en faktor som forekommer i for eksempel  $z_j$  vil dermed i første omgang endre generøsiteten. Men denne enringen igjen, har en direkte konsekvens for lønnsulikheten, som deretter virker tilbake på generøsiteten. Det er dette som er multiplikatoreffekten - *The Equality Multiplier*. Hovedfokuset i deres artikkel ligger i estimering av  $a_I$  og  $a_g$ , som viser henholdsvis elasticiteten  $G_j$  med hensyn på  $I_j$  og elasticiteten i  $I_j$  med hensyn på  $G_j$ .

Systemets likevektsverdier for konstante eksogene kovariater  $A_j$  og  $B_j$ , er:

$$\begin{aligned} \ln(G_j) &= m [A_j - a_I B_j] \\ \ln(I_j) &= m [B_j - a_g A_j] \end{aligned}$$

hvor  $m = \frac{1}{1-a_I a_g} > 1$ , så sant  $a_I a_g < 1$ , er likhetsmultiplikatoren.

Tallmaterialet som ligger til grunn for utregningen av disse størrelsene,  $a_I$ ,  $a_g$  og  $m$ , er OECD sin inntekstdatabase (18 land fra 1976 - 2002) samt *Comparative Welfare Entitlement Dataset* som er beregnet og tilgjengeliggjort av Lyle Scruggs ved University of Connecticut, som gir tall for velferdsstatsgenerøsitet.

Slik som modelleringen er nå, går det ikke an å få konsistente resultater ved hjelp av vanlig minste kvadraters metode (OLS), fordi den forutsetter at høyresidevariablene er eksogene for den avhengige variabelen. Slik som det er modellert her er  $\ln(G_j)$  og  $\ln(I_j)$  spesifikt *endogene* for hverandre. Dermed blir  $a_I$  og  $a_g$  ved OLS konsekvent over- eller underestimert, uansett antall observasjoner. I deres analyse blir dette løst ved å kjøre en 3SLS-analyse<sup>3</sup> på datamaterialet. Endogenitetsproblemet blir altså løst ved å bruke instru-

---

<sup>2</sup>en viss overlapping med  $z_j$  forekommer

<sup>3</sup>3SLS er kombinasjonen av 2SLS og SUR

mentvariable.

Instrumentvariable for lønnsulikhet er:

- Grad av koordinasjon i lønnsforhandlinger
- Arbeidere i arbeidskonflikt
- Andel av befolkningen med høyere utdanning
- Sysselsettingsgrad i 16-64-befolkningen.

Instrumentvariable for velferdsstatsgenerøsitet er:

- Høyrepartistyrte regjering
- trend

Begge modellene inkluderer faste virkninger for land.

Empirien bekrefter teorien; de viktigste resultatene er, (med standardfeil i parentes):

$$a_I = -.64 \text{ (.1251)}$$

$$a_g = -.53 \text{ (.0744)}$$

$$m = 1.52$$

Et eksogent sjokk som direkte påvirker enten lønnsulikheten eller velferdsstatsgenerøsiteten, vil, etter at de interagerende effektene har stabilisert seg, resultere i en økning i den relevante størrelsen på 152 prosent av det initialte sjokket. Det er likhetsmultiplikatoreffekten.

## 3.2 Grunnlaget for alternativ modellering

Tore Schweder, Trond Reitan og Jorijntne Henderiks gjør i artikkelen *Modelling Evolution by Linear Stochastic Differential Equations with Latent Layers* [Reitan, Schweder, Henderiks, 2009] et studie av utviklingen i algestørrelse over 63 millioner år. Det umiddelbare spørsmålet er åpenbart; hva gjør en slik referanse i en mastergradsoppgave i økonomi? Jeg skal vedgå at det er lettere å finne forskjeller mellom alger og inntekstulikhet og velferdsstatsgenerøsitet, enn likheter, men felles for dem er likevel det at de er størrelser

i utvikling. Schweder, Reitan og Henderiks begrunner algeutviklingen med naturlig utvalg, altså darwinistisk utvikling. Velferdsstatsgenerøsitet og inntektsulikhet er også størrelser i utvikling. Politiske vedtak, grasrotbevegelser, fagforeningsarbeid, samfunnsstruktur i det store og hele, internasjonale politiske begivenheter og ikke minst global økonomisk vekst er faktorer som bidrar til utviklingen av inntektsulikhet og velferdsstatsgenerøsitet. Det som er felles er altså *at* de utvikler seg, og *ikke* begrunnelsen i seg selv. Et rammeverk som egner seg for *en* utviklingsmodell, kan kanskje, etter passende remodelleringer, egne seg for en *annen* utviklingsmodell.

Utvikling er endring *over tid*, og mange interessante spørsmål vedrørende et fenomen, som for eksempel velferdsstatsgenerøsitet, inkluderer en tidsdimensjon. Hvis vi tenker oss at en prosess kan modelleres som bevegelse omkring en trend, så kan det være interessant å vite hvor lang tid det tar å “hente inn” igjen trenden, hvis en på et tidspunkt ligger under. Slike spørsmål sier noe om “tregheten” i “systemet”. Ved å se på samme aspekt for forskjellige politiske tradisjoner, kan vi bestemme grad av flyktighet, eller tyngde eller soliditet i systemet, og gjøre sammenligninger.

Differensialligninger er beskrivelser av utvikling. Hvis man kjenner en prosessstilstand på et tidspunkt, og kjenner beskrivelsen av utviklingen, altså bevegelsen, så kan man bestemme hele prosessens bane i tiden. De er altså deterministiske i natur. Ved å legge til et usikkerhetselement, et stokastisk bidrag, så gir man utviklingsbanen en sannsynlighetsfordeling. De prosesser vi gjerne ønsker å modellere innen økonomifaget er gjerne preget av iboende usikkerhet, og modellvalget bør reflektere denne siden av virkeligheten. Samtidig har man en viss forutsigbarhet i størrelser hvor kulturelle normer er av betydning. Det er for eksempel lite sannsynlig av gjennomsnittslønnen i Norge vil falle til det halve innen kort tid. Stokastiske differensialligninger kombinerer det sikre med det usikre, og kan dermed brukes til modellering av situasjoner styrt av menneskelig aktivitet, hvor vi har nettopp det stabile og det uforutsigbare samtidig. Størrelser som velferdsstatsgenerøsitet og inntektsulikhet er påvirket av den kulturen de observeres fra, men er også gjenstand for påvirkning fra flyktige økonomiske eller politiske begivenheter.

En stokastisk differensialligning som er hensiktsmessig i slik modellering

er Ornstein-Uhlenbeck-prosessen (OU). Den er karakterisert ved to parametre: En diffusjonsparameter og en drift-parameter. Drift-parameteren er det som bestemmer styrken på det deterministiske elementet, mens diffusjonsparameteren bestemmer styrken til det stokastiske elementet.

En gradvis oppbygging av OU-prosessen fra en vanlig lineær differensial-ligning er lagt til første del av appendixet, og kan være hensiktsmessig lesning for å komme inn i tankegangen.

Den grunnleggende Ornstein-Uhlenbeck-prosessen ser slik ut:

$$dX(t) = -\alpha(X(t) - \mu)dt + \sigma dW(t)$$

Vi forutsetter her at  $\alpha > 0$ , for å få stasjonaritet. Det deterministiske elementet viser hvor sterkt  $X(t)$  trekkes mot sitt likevektsnivå,  $\mu$ .  $\alpha$  er prosessens drift-paramter, og trekkstyrken er avhengig av avstanden mellom  $X(t)$  og  $\mu$ .  $dW(t)$  er differensialet til en standardisert Wiener-prosess<sup>4</sup> og representerer de kontinuerlige normalfordelte forstyrrelsene i veksthastigheten. Disse blir multiplisert med skaleringen  $\sigma$ , som er diffusjonsparameteren. En høy diffusjonsparameter betyr at stokastikken får stor forklaringsvekt, og motsatt.

Hvis vi ønsker å bruke denne ligningen i en økonomisk modell må vi løse det umiddelbare problemet: Hvilke økonomiske størrelser kan tenkes å ha en utviklingsbane som ligner på det som her skisseres?, altså kontinuerlig bevegelse mot et visst likevektspunkt, men med stadige små forstyrrelser? Slike ligninger har vært brukt til prisutvikling i aksjemarkedet, hvor en aksje kan tenkes å ha en fundamentalt begrunnet verdi, her representert ved  $\mu$ . Aksjeprisen flytter seg kontinuerlig omkring denne verdien på grunn av stadige kjøp og salg. Av og til er prisen over, andre ganger under, men oftest i nabolaget av den fundamentale verdien. Denne forklaringen egner seg dog bedre som forklaring på OU-prosessen, enn på aksjeprising. Et annet bidrag til intuisjonen finner vi fra [Reitan, Schweder, Henderiks, 2009] hvor  $\mu$ , her noe forenklet, modelleres som den optimale algestørrelsen, gitt miljøet.  $\mu$  representerer målet for utviklingen.

---

<sup>4</sup>Se appendix for beskrivelse

Hvis vi ønsker å modellere inntekstulikheten på denne måten, så gjør vi implisitt en antagelse om at inntekstulikheten har konstant forventning. Når vi observerer en slik størrelse over tid, ser vi at det ikke er en gunstig tolkning, og at denne én-dimensjonale modellen ikke er godt egnet til formålet.

En mer realistisk tilnærming kan vi få ved å lage et system av OU-prosesser. Den første ligningen trekkes da ikke mot en konstant verdi, men mot nivået til den andre prosessen. Den andre prosessen trekkes mot en konstant. Det beskrives da som en to-lags-modell. Det øverste laget følger det underste laget, som igjen følger en konstant. Litt enklere forklart betyr det at  $\mu$ , slik jeg her har brukt den, *selv modelleres som en Ornstein-Uhlenbeck-prosess*.

### 3.3 Ligningssystemer

I ligningssystemet

$$\begin{aligned}dX_1(t) &= -\alpha_1(X_1(t) - X_2(t))dt + \sigma_1 dW_1(t) \\dX_2(t) &= -\alpha_2(X_2(t) - \mu)dt + \sigma_2 dW_2(t)\end{aligned}$$

hvor  $W_1$  og  $W_2$  er to uavhengige Wiener-prosesser, ser vi at  $X_1$  følger  $X_2$ , og at  $X_2$  fluktuerer omkring konstanten  $\mu$ , gitt at  $\alpha_1, \alpha_2 > 0$ . Det øverste laget er beskrevet av  $X_1$ , mens det underliggende laget er beskrevet av  $X_2$ . Kausalstrukturen er  $X_2 \rightarrow X_1$ . En lettere modifisert versjon av denne modellen er brukt konkret i oppgaven, og benevnes som *Modell 1*.

Den førstnevnte ligningen har to parametre, og har dermed en viss fleksibilitet. Ved å se for oss at samme datamateriale er generert av en slik to-lags modell, som her er beskrevet, hvor det øverste laget direkte generer de observerte data, men at denne prosessen er *påvirket* av den underliggende prosessen, på den måten som er modellert, så får vi dobbel stokastikk, samt to nye parametre som kan forklare observasjonene. Språkbruken her er verdt å presisere:  $X_1(t)$  sies å representere *topplaget*, og  $X_2(t)$  representerer det *underliggende* laget, fordi denne påvirker  $X_1(t)$ . En slik modell er mer fleksibel, og kan bedre treffe observasjonene. Tankegangen her er at prosessen modellert som  $X_2$  ikke er observert direkte, men at dens virkning kun kan

påvises gjennom den virkning den har på  $X_1$ . Med dobbel stokastikk menes det direkte stokastiske elementet som hører til topplaget,  $dW_1(t)$ , og topplagets filtrering av det underliggende lagets stokastikk, altså  $dW_2$ . Dette blir spesifisert i utregningene under. Igjen vil jeg spesifisere tydelig at denne to-lags-modellen ikke forutsetter flere observasjoner enn det som var forutsatt for den én-dimensjonale modellen. Begge modellene skal forklare det samme observerte datamaterialet, mens i den siste modellen er  $X_2(t)$  bragt inn som et usynlig underliggende lag som forklarer topplagsprosessen, men som ikke genererer datamateriale adskilt fra det vi har fra  $X_1(t)$ . Begrepene topplag og bunnlag går igjen i resten av oppgaven.

Det er viktig å spesifisere her at målet med oppgaven ikke er å finne en modellering som best forklarer dataene, *med mindre* elementene i modellen kan gi økonomisk tolkning. Det er mulig å legge til så mange lag at alle observasjonene godt kan forklares. Det blir etter hvert umulig å gi noen fornuftig tolkning av de forskjellige lagene, og modellen mister dermed sin hensiktsmessighet.

For å tydeliggjøre hva som kan være potensielle tolkninger av lagene, så fortsetter jeg med to-lags-modellen slik den er spesifisert over. Hvis vi bruker denne modellen til å forklare inntekstulikhet, for eksempel, så er  $X_1$  nettopp inntekstulikhetsprosessen.  $X_1$  følger  $X_2$ , så disse må ha samme natur. Altså må  $X_2$  også være en inntekstulikhetsprosess. Men hvilken, hvis  $X_1$  allerede representerer den observerte inntekstulikheten? Inntekstulikhet er ikke en prosess i fri bevegelse, som en random walk. Den er styrt av de politiske tiltak som gjøres, og styrket eller begrenset av den økonomiske realiteten. Den er et resultat av handling og vilje. Det som til enhver tid er inntekstulikheten i samfunnet er en størrelse i endring. Den er alltid på vei mot noe. Poenget her er ikke at den er på vei mot det som er uttalt politisk, men til enhver tid trekker den i en retning. Den underliggende prosessen,  $X_2$ , kan dermed tenkes å representere denne “samfunnsambisjonen”. En noe vag definisjon kan være greit nok på en usynlig prosess.

Vi ser på to størrelser, inntekstulikhet,  $I(t)$ , og velferdsstatsgenerøsitet,

$G(t)$ , og disse interagerer:

$$dI(t) = -\alpha_I(I(t) - \mu_I + a_G(G(t) - \mu_G))dt + \sigma_I dW_I(t)$$

$$dG(t) = -\alpha_G(G(t) - \mu_G + a_I(I(t) - \mu_I))dt + \sigma_G dW_G(t)$$

Jeg forklarer den første ligningen, da de er symmetriske. I en situasjon hvor  $G(t) = \mu_G$ , så ser vi at  $I(t)$  *synker*, hvis den ligger over sin forventningsverdi, og motsatt hvis den ligger under. Hvis vi har  $I(t) = \mu_I$ , så vil  $I(t)$  synke hvis  $G(t)$  ligger over sin forventningsverdi,  $\mu_G$ , og synke hvis den ligger under. I mellom disse spesialsituasjonene, altså med verdier lik sin forventnings som utgangspunkt, så vil dene være de samme, men ikke den absolutte retningen. For eksempel vil  $I(t)$  kunne være *over* sin forventning, som i seg selv bidrar til at  $I(t)$  *synker*, mens  $G(t)$  kan være *under* sin forventning, som i seg selv vil bidra til at  $I(t)$  *stiger*. Det kan enten føre til at  $I(t)$  synker i mindre grad, eller at den i sum stiger, avhengig av hvilken effekt som er sterkest. Vi søker verdiene på de parametrene,  $\alpha_I$ ,  $\alpha_G$ ,  $\mu_I$ ,  $\mu_G$ ,  $a_I$ ,  $a_G$ ,  $\sigma_I$  og  $\sigma_G$  som *best* forklarer den observerte tidsutvikling for nettopp inntektsulikhet og velferdsstatsgenerøsitet for et land. Denne spesifikke modellformuleringen med tilhørende forklaring forutsetter at interaksjonsparametrene,  $a_I$  og  $a_G$  begge er positive. Denne modellen kan synes enkel, men det er fordi den kun gjelder *ett* land. Vi skal se på flere land, og det er ikke naturlig å tro at alle land passer inn i nettopp dette rammeverket. Hvis vi skal se på globale interaksjonsparametre, er det naturlig å holde disse like i en modell som omfatter flere land, men drift- og diffusjonsparametrene kan være forskjellige, og iallefall er det sikkert at forventningsverdiene,  $\mu_I$  og  $\mu_G$ , er forskjellige. Dette medfører at antall parametre øker når flere land inkluderes i modellen. Dette er lite hensiktsmessig og gjør estimeringen heller vanskelig. Det er bedre å estimere få parametre basert på flere observasjoner, enn å la parametrene øke i antall i takt med observasjonene. Dette blir løst ved å transformere datamaterialet, slik at landsspesifikke egenskaper blir fjernet. Dermed blir det mulig å øke antall land i modellen, men likevel holde parameterrommet begrenset. Jeg kommer tilbake til dette i de spesifikke modellformuleringene. Det er en oppgave som inkluderer mye matematikk. Systemer av stokastiske

differentielligninger omskrives til matriseform, og kovariansmatriser utregnes. Et enkelt tilfelle blir regnet ut i sin helhet, og en variant av denne utregningen blir brukt i en modell senere.

### 3.4 Sannsynlighetsmaksimering

En kortfattet beskrivelse av sannsynlighetsmaksimeringsteorien er lagt til appendixet. Dette er kun en intuitiv fremstilling. Når vi har én tidsserie og ser for oss at denne er generert av en fler-parameter-modell, så kan vi ikke finne analytiske uttrykk for parameterne. Mange kombinasjoner av parametre kan (normalt sett) tenkes å genere det observerte utfallet. Spørsmålet er hvilken av disse kombinasjonene er den rette? I teorien er svaret enkelt; vi søker  $\theta_m$  og:  $\theta_m = \operatorname{argmax}_{\theta} \operatorname{Lik}(\theta | x)$ , hvor  $\theta$  er vektoren av parametre,  $x$  er vektoren av observasjoner, og  $\operatorname{Lik}(\theta|x)$  er *likelihood-funksjonen*. Vektorer av observasjoner fra en Ornstein-Uhlenbeck-prosess er multinormalfordelte. Forventningsverdien er konstant i de stasjonære modellen, og lik  $\mu \in R$ , og sammenhengen mellom alle observasjonene er beskrevet av kovariansmatrisen. Siden prosessen forventes å trekke gradvis mot  $\mu$ , vil vi få forventet sammenheng mellom to forskjellige observasjoner. Hvis et stokastisk element på et tidspunkt  $t$  trekker prosessen opp, vil prosessen være forventet, etter  $t$ , å trekke gradvis mot  $\mu$ , men likevel være høyere enn hvis det nevnte stokastiske innslaget *ikke* hadde kommet. Slik er sammenhengen mellom alle observasjonene bundet sammen av kovariansmatrisen, og denne kovariansmatrisen er igjen en funksjon av de parametrene som inngår i modellen. En modell med høy verdi på  $\alpha$ , som betyr at prosessen raskt trekkes mot forventningen, vil ha lavere korrelasjon mellom observasjonene, fordi den har kortere “hukommelse”.

Modellens sannsynlighetsfordeling er altså beskrevet av alle dens parametre. For en realisert vektor av observasjoner, hvor vi ønsker å bruke modellen, vil forskjellige sett parameterverdier, altså  $\theta$ -verdier, gi modellen forskjellig forklaringssevne. Hvis vi observerer at mange observasjoner ligger under forventningen i flere tidsperioder på rad, er det naturlig å tro at en lav  $\alpha$ -verdi vil være hensiktsmessig. Hvis utviklingen er veldig jevn, altså lite variasjon



i observasjonene, er det naturlig å tillegge stokastikk liten verdi, og således falle på lave  $\sigma$ -verdier.

Som nevnt ovenfor er det teoretisk enkelt å bestemme parametervektoren  $\theta_m$ , iallefall når likelihood-funksjonen er éntoppet. I praksis er det langt vanskeligere. En potensiell framgangsmåte er å avgrense et fornuftig intervall for hver av de  $n$  parametrene, og dermed gjennomløpe en  $n$ -dimensjonal boks, etter en passende intervalloppdeling. En datamaskin kan instrueres om å gjøre dette, men det vil være enormt ressurskrevende, og må utsettes til fremtidens kraftigere datamaskiner. Det finnes imidlertid algoritmer for å finne optimum. Den algoritmen som jeg bruker her har en annen framgangsmåte. Et tilfeldig  $n$ -dimensjonalt punkt blir valgt, og likelihooden blir beregnet. Algoritmen gjør numerisk derivasjon ved å bevege seg litt i det  $n$ -dimensjonale rommet, og måle likelihooden igjen. Hvis likelihooden er høyere, var retningsvalget riktig, og samme prosedyre gjentas til likelihooden ikke endres (nok) lenger. Det er derimot *ikke* gitt at dette topp-punktet er det globale topp-punktet. For å overkomme dette problemet (etter beste evne), blir “hill climbingen”<sup>5</sup> repetert flere ganger, hver gang med et nytt tilfeldig utgangspunktet i parameterrommet. Håpet er dermed å utpeke en parametervektor en kan tro på.

### 3.5 Likelihood-funksjonen

Observasjonene vi har er spredt i tid, og vi har et litt hullede tallmateriale. Det er alltid en ulempe å mangle datamateriale *per se*, men i dette tilfellet bidrar det til å vise en av modellens fordeler. Den stiller nemlig ingen krav til hvor ofte eller spesifikt *når* observasjonene skal forekomme. Det viktige er derimot å kunne tidfeste observasjonene, og i nettopp den situasjonen er vi.

Denne tankegangen er grunnleggende i denne oppgaven. Jeg modellerer prosessene som flerlags OU-prosesser, og setter dem inn i større simultane systemer. Hver prosess er altså en Gauss-prosess, og er dermed beskrevet av forventningen og kovariansen. Høyrere ordens momenter bringer ikke informasjon. I likelihoodberegningen er det nettopp denne egenskapen som

---

<sup>5</sup>minner om Newtons metode

er viktig. Observasjoner av en prosess på to forskjellige tidspunkt har en bestemt kovarians og bestemte forventningsverdier. Disse er en funksjon av modellens parametre. Alle observasjoner vil ha et strukturelt forhold til alle andre observasjoner. Ved  $n$  observasjoner får vi altså en kovariansmatrise med dimensjon  $n \times n$ . I likelihooden er det disse som ligger til grunn for optimeringen. Derfor vil jeg først integrere OU-prosessen, for deretter å vise forventning og kovarians. I denne delen av oppgaven kommer mange lange utregninger. Disse er med for kompletthets skyld, men er ikke viktige i seg selv. Det er kun resultatene som brukes. Jeg legger en utregning av varians, kovarians og forventning av en én-dimensjonal prosess til appendikset, da all tankegang som ligger her, også inngår i de mer kompliserte utregningene som følger av fler-dimensjonale prosesser. Fra nå av forenkles disse ved hjelp av matrise-notasjon..

To-lags-modellen brukt ovenfor kan skrives i matrise-notasjon, slik det gjøres i [Schweder, Reitan, Henderiks, 2009].

$$dX(t) = (m(t) + AX(t))dt + \Sigma dW(t)$$

hvor

$$\begin{aligned} dX(t) &= \begin{pmatrix} dX_1(t) \\ dX_2(t) \end{pmatrix} & X(t) &= \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} & m(t) &= \begin{pmatrix} 0 \\ \alpha_2 \mu \end{pmatrix} \\ A &= \begin{pmatrix} -\alpha_1 & \alpha_1 \\ 0 & -\alpha_2 \end{pmatrix} & \Sigma &= \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} & dt &= \begin{pmatrix} dt \\ dt \end{pmatrix} \\ dW(t) &= \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix} \end{aligned}$$

Dette er altså et ligningssystem i med to ligninger, men samme notasjon kan generaliseres til et vilkårlig antall ligninger.

Komponentene og koeffisientene vil i det generelle tilfellet ha følgende

dimensjonalitet:

$$\begin{array}{ll} dX(t) - X(t) \, dt - m(t) & (n \times 1) \\ A & (n \times n) \\ dW(t) & (m \times 1) \\ \Sigma & (n \times m) \end{array}$$

Her er det verdt å merke seg at vi generelt har  $n \neq m$ , men alltid  $m \leq n$ . Vi vil i det normale tilfellet likevel ha  $m = n$ , med positive elementer kun på hoveddiagonalen til  $\Sigma$ . Det betyr enkelt og greit at hver ligning i systemet får sitt egne stokastiske bidrag. Utrekningene i det kommende vil basere seg på det, men det kan også være aktuelt i spesielle modelleringer å gi flere ligninger det samme stokastiske bidraget. Det vil i de tilfellene begrunnes på økonomisk grunnlag, og ha sitt opphav i hva vi ønsker å undersøke.

Ifølge regel 19.30 i *Matematisk formelsamling for økonomer* har eksponensialmatrise følgende definisjon:

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$$

hvor  $A$  er en  $n \times n$ -matrise. Den tilsvarende definisjonen for den én-dimensjonale eksponensialformelen er:

$$e^a = \sum_{n=0}^{\infty} \frac{1}{n!} a^n$$

hvor  $a$  er en vilkårlig konstant skalar.

Regel 19.31 (ibid.) gir oss to egenskaper vi vil bruke i den videre utviklingen av den generaliserte modellen:

$$(e^A)^{-1} = e^{-A} \quad \frac{d}{dx}(e^{xA}) = Ae^{xA}$$

hvor  $x \in \mathbb{R}$  og utfra definisjonen er det tydelig at  $e^A x$  kommuterer:  $Ae^{Ax} = e^{Ax}A$ .

Ved å integrere matrisesystemet finner vi dets Itô-representasjon, som i

[Reitan, Schweder, Henderiks 2009]:

$$\begin{aligned}
dX(t) &= (m(t) + AX(t))dt + \Sigma dW(t) \\
dX(t) - AX(t)dt &= m(t)dt + \Sigma dW(t) \\
e^{-At}dX(t) - e^{-At}AX(t)dt &= e^{-At}m(t)dt + e^{-At}\Sigma dW(t) \\
e^{-At}dX(t) - Ae^{-At}X(t)dt &= e^{-At}m(t)dt + e^{-At}\Sigma dW(t) \\
\frac{d}{dt}[e^{-At}X(t)] &= e^{-At}m(t)dt + e^{-At}\Sigma dW(t) \\
e^{-At}X(t) &= X(0) + \int_0^t e^{-As}m(s)ds + \int_0^t e^{-As}\Sigma dW(s) \\
X(t) &= e^{At}X(0) + e^{At} \int_0^t e^{-As}m(s)ds + e^{At} \int_0^t e^{-As}\Sigma dW(s) \\
X(t) &= e^{At}X(0) + \int_0^t e^{A(t-s)}m(s)ds + \int_0^t e^{A(t-s)}\Sigma dW(s)
\end{aligned}$$

Hvor det kan nevnes at integralet til en matrise, er matrisen av integra-lene:

$$\int \begin{pmatrix} a & b \\ c & d \end{pmatrix} dt = \begin{pmatrix} \int a dt & \int b dt \\ \int c dt & \int d dt \end{pmatrix}$$

Hvis vi skal regne ut dette i sin nåværende form, innebærer det å gjen-nomføre uendelig mange regneoperasjoner, fordi vi trenger eksponensialma-trisen, hvis definisjon kan gjentas:  $e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$ . Det er åpenbart umulig i den formen, for en generell  $A$ -matrise, men vi løser problemet ved hjelp av egenverdidekomponering: Fra *Lineær Algebra* [Sydsaeter, 2006b, 199] har vi følgende:

“Generelt sier vi at en  $(n \times n)$  matrise  $A$  er *diagonaliserbar* dersom det finnes en invertibel  $(n \times n)$  matrise  $P$  og en diagonalmatrise  $D$ , slik at  $P^{-1}AP = D$ .” Ved multiplikasjon fra venstre med  $P$  og fra høyre med  $P^{-1}$  får vi  $A = PDP^{-1}$ , og ved induksjon kan en vise at  $A^n = PD^nP^{-1}$ :

$$A^{n+1} = A^n A = PD^n P^{-1} P D P^{-1} = PD^n D P^{-1} = PD^{n+1} P^{-1}$$

og det vises nedenfor at dette stemmer for  $n = 1$ . At det stemmer for  $n = -1$  blir gjort bruk av senere.

Diagonalmatriser er enklere å regne med: Hvis

$$D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$

så har vi

$$D^n = \begin{pmatrix} d_1^n & 0 & 0 \\ 0 & d_2^n & 0 \\ 0 & 0 & d_3^n \end{pmatrix}$$

og resultatet generaliseres til alle diagonalmatriser.

Noen spesielle matriser kan alltid diagonaliseres, ifølge setning 10.2 i *Lineær Algebra*: Anta at  $A$  er en  $n \times n$  matrise med  $n$  forskjellige egenverdier. Da er  $A$  diagonaliserbar.

Vi forutsetter i det videre at de  $A$ -matrisene vi jobber med har denne egenskapen, og det vil bli begrunnet med et økonomisk resonnement når den grunnleggende modellen blir skissert.

Dette bruker vi i utregningen av

$$\begin{aligned}
e^{At} &= \sum_{n=0}^{\infty} \frac{1}{n!} (At)^n \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} (PDP^{-1}t)^n \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} PD^n P^{-1} t^n \\
&= P \sum_{n=0}^{\infty} \frac{1}{n!} D^n t^n P^{-1} \\
&= P \begin{pmatrix} \sum_{n=0}^{\infty} \frac{1}{n!} d_1^n t^n & 0 & \dots & 0 \\ 0 & \sum_{n=0}^{\infty} \frac{1}{n!} d_2^n t^n & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sum_{n=0}^{\infty} \frac{1}{n!} d_n^n t^n \end{pmatrix} P^{-1} \\
&= P \begin{pmatrix} e^{-d_1 t} & 0 & \dots & 0 \\ 0 & e^{-d_2 t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{-d_n t} \end{pmatrix} P^{-1} \\
&= P e^{Dt} P^{-1}
\end{aligned}$$

Setning 10.1: En  $n \times n$  matrise  $A$  er diagonaliserbar hvis og bare hvis  $A$  har  $n$  lineært uavhengige egenvektorer  $x_1, x_2, \dots, x_n$ . I så fall er

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

der  $P$  er matrisen med  $x_1, x_2, \dots, x_n$  som kolonner og  $\lambda_1, \dots, \lambda_n$  er de tilhørende egenverdiene.

Vi forutsetter at  $P$  er matrisen bestående av egenvektorene til  $A$ , og at  $D$  er diagonalmatrisen med de tilhørende egenverdiene.

Hvis  $A = PDP^{-1}$ , så er  $A^{-1} = PD^{-1}P^{-1}$ , som lett kan vises:

$$AA^{-1} = PDP^{-1}PD^{-1}P^{-1} = I$$

Vi erstatter  $e^{At}$  i den første Ito-representasjonen, med  $Pe^{Dt}P^{-1}$ :

$$X(t) = Pe^{Dt}P^{-1} \left[ X(0) + \int_0^t Pe^{-Du}P^{-1}m(u)du + \int_0^t Pe^{-Du}P^{-1}\Sigma dW(u) \right]$$

eller, ved utregning:

$$\begin{aligned} X(t) &= Pe^{Dt}P^{-1}X(0) + Pe^{Dt}P^{-1} \int_0^t Pe^{-Du}P^{-1}m(u)du + Pe^{Dt}P^{-1} \int_0^t Pe^{-Du}P^{-1}\Sigma dW(u) \\ &= Pe^{Dt}P^{-1}X(0) + \int_0^t Pe^{D(t-u)}P^{-1}m(u)du + \int_0^t Pe^{D(t-u)}P^{-1}\Sigma dW(u) \end{aligned}$$

Ved hjelp av egenvektordekomponeringen har vi altså funnet en grei måte å representere systemet på, men det er veldig lite intuitivt.  $A$ -matrisen inneholder all informasjon om systemets interaksjon. Generelt vil én ligning være avhengig av en annen, som igjen er avhengig av en tredje, og så videre. Det er denne avhengigheten som gjør det vanskelig å regne ut. Det som skjer i egenvektordekomponeringen er at de retningene hvor systemet reproduseres (i en viss størrelsesorden), isoleres ut og gjøres til basevektorer for et nytt koordinatsystem. I dette nye koordinatsystemet er hver ligning uavhengig av de andre, representert ved *diagonal*matrisen, og dette gjør det mulig i praksis å regne ut det integrerte systemet. Den inverse matrisen til det nye koordinatsystemet sørger for å transformere rommet tilbake til dét vi observerer størrelsene i. Ved å gå via en annen base kan vi altså regne ut og finne størrelser for systemet.

Systemet beskriver en *Gauss-prosess*, og det har følgende første og andre moment:

$$E[X(t)] = Pe^{Dt}P^{-1}X(0) + \int_0^t Pe^{D(t-u)}P^{-1}m(u)du$$

fordi forventningene til de stokastiske elementene er hver for seg lik 0, og

således også for deres integral. Kovariansen er:

$$\begin{aligned}
Cov[X(v), X(t)] &= E[(X(v) - E[X(v)])(X(t) - E[X(t)])^T] \\
&= E[(\int_0^v Pe^{D(v-u)} P^{-1} \Sigma dW(u))(\int_0^t Pe^{D(t-u)} P^{-1} \Sigma dW(u))^T] \\
&= E[(\int_0^v Pe^{D(v-u)} P^{-1} \Sigma dW(u))(\int_0^v Pe^{D(t-u)} P^{-1} \Sigma dW(u))^T] \\
&= \int_0^v Pe^{D(v-u)} P^{-1} \Sigma^2 (P^{-1})^T e^{D(t-u)} P^T du
\end{aligned}$$

hvor vi har gjort bruk av følgende:

- $(ABC)^T = C^T B^T A^T$
- $\int_a^c dt = \int_a^b dt + \int_b^c dt$
- $E[(\int_0^v Pe^{D(v-u)} P^{-1} \Sigma dW(u))(\int_v^t Pe^{D(t-u)} P^{-1} \Sigma dW(u))^T] = 0$
- repetitiv bruk av Itô-isometrien [Seierstad, 2009]:  $E[\int_0^v g(t) dW(t)]^2 = \int_0^t (g(t))^2 dt$

For et mer konkret eksempel går vi tilbake til den første to-lags-modellen, som brukes konkret på datamaterialet, og hvor resultatet er å finne i appendix 8.7:

$$\begin{aligned}
dX_1(t) &= -a_1(X_1(t) - X_2(t))dt + \sigma_1 dW_1(t) \\
dX_2(t) &= -a_2(X_2(t) - \mu)dt + \sigma_2 dW_2(t)
\end{aligned}$$

som på matrise-form hadde følgende komponenter:

$$\begin{aligned}
dX(t) &= \begin{pmatrix} dX_1(t) \\ dX_2(t) \end{pmatrix} & X(t) &= \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} & m(t) &= \begin{pmatrix} 0 \\ \alpha_2 \mu \end{pmatrix} \\
A &= \begin{pmatrix} -\alpha_1 & \alpha_1 \\ 0 & -\alpha_2 \end{pmatrix} & \Sigma &= \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} & dt &= \begin{pmatrix} dt \\ dt \end{pmatrix} \\
dW(t) &= \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}
\end{aligned}$$

Vi finner  $A$ -matrisens egenverdier og tilhørende egenvektorer. Matrisen  $P$  ved å sette sammen egenvektorene som kolonner i  $A$ , altså på vanlig måte.



Eigenverdimatrisen blir en diagonalmatrise med eigenverdiene i samme kolonneposisjon som dens tilhørende egenvektor har i  $P$ -matrisen.

$$Ax = \alpha x$$

$$\Leftrightarrow (A - \alpha I)x = 0$$

Ifølge [Sydsaeter, 2006b, setning 6.5, avsnitt 6.3] har dette homogene lineære likningssystemet en løsning  $x \neq 0$ , hvis og bare hvis systemets koeffisientmatrise har determinant lik 0, altså hvis:

$$|A - \alpha I| = 0$$

Siden matrisen  $A - \alpha I$  har nullverdier nedenfor hoveddiagonalen, så er eigenverdiene lik elementene på hoveddiagonalen, altså  $-\alpha_1$  og  $-\alpha_2$ .

For eigenverdien  $-\alpha_1$  blir ligningssystemet lik:

$$0v_1 + \alpha_1 v_2 = 0$$

$$0v_1 + (\alpha_1 - \alpha_2)v_2 = 0$$

som gir oss egenvektoren

$$V = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \propto \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

For eigenverdien  $-\alpha_2$  blir ligningssystemet lik:

$$\begin{cases} (-\alpha_1 + \alpha_2)v_1 + \alpha_1 v_2 = 0 \\ 0v_1 + 0v_2 = 0 \end{cases}$$

$$\Leftrightarrow v_1 + \frac{\alpha_1}{\alpha_2 - \alpha_1} v_2 = 0$$

som gir oss egenvektoren

$$V \propto \begin{pmatrix} \frac{-\alpha_1}{\alpha_2 - \alpha_1} \\ 1 \end{pmatrix}$$

Vi har således funnet vektoren

$$P = \begin{pmatrix} 1 & \frac{-\alpha_1}{\alpha_2 - \alpha_1} \\ 0 & 1 \end{pmatrix}$$

hvis inverse er gitt av *Lineær algebra*[Sydsaeter, 2006b, regel 6.3], som sier:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow M^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Det følger umiddelbart at

$$P^{-1} = \begin{pmatrix} 1 & \frac{\alpha_1}{\alpha_2 - \alpha_1} \\ 0 & 1 \end{pmatrix}$$

og ved bruk av egenverdimatrisen som beskrevet ovenfor, får vi:

$$A = \begin{pmatrix} -\alpha_1 & \alpha_1 \\ 0 & -\alpha_2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{-\alpha_1}{\alpha_2 - \alpha_1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -\alpha_1 & 0 \\ 0 & -\alpha_2 \end{pmatrix} \begin{pmatrix} 1 & \frac{\alpha_1}{\alpha_2 - \alpha_1} \\ 0 & 1 \end{pmatrix}$$

som bekreftes gjennom direkte utregning, men er utelatt her.

Samlet sett har vi nå alt vi trenger til å regne ut prosesslagenes utvikling. Vi begynner med det underliggende laget, beskrevet av  $X_2(t)$ , som altså er en én-dimensjonal Ornstein-Uhlenbeck-prosess, før vi regner ut utviklingen av  $X_1(t)$ , som altså filtrerer den underliggende prosessen: Vi tar utgangspunkt i:

$$X(t) = Pe^{Dt}P^{-1}X(0) + \int_0^t Pe^{D(t-u)}P^{-1}m(u)du + \int_0^t Pe^{D(t-u)}P^{-1}\Sigma dW(u)$$

Siden alle matrisene er spesifisert ovenfor, blir ingen av utregningene her vist. Resultatene er:

$$Pe^{Dt}P^{-1} = \begin{pmatrix} e^{-\alpha_1 t} & \frac{\alpha_1}{\alpha_2 - \alpha_1}(e^{-\alpha_1 t} - e^{-\alpha_2 t}) \\ 0 & e^{-\alpha_2 t} \end{pmatrix}$$

$$Pe^{D(t-u)}P^{-1}m = \begin{pmatrix} \alpha_2\mu\frac{\alpha_1}{\alpha_2-\alpha_1}(e^{-\alpha_1(t-u)} - e^{-\alpha_2(t-u)}) \\ \alpha_2\mu e^{-\alpha_2(t-u)} \end{pmatrix}$$

$$Pe^{D(t-u)}P^{-1}\Sigma = \begin{pmatrix} \sigma_1 e^{-\alpha_1(t-u)} & \sigma_2\frac{\alpha_1}{\alpha_2-\alpha_1}(e^{-\alpha_1(t-u)} - e^{-\alpha_2(t-u)}) \\ 0 & \sigma_2 e^{-\alpha_2(t-u)} \end{pmatrix}$$

Av dette trekker vi at første komponent,  $X_1(t)$  har følgende beskrivelse:

$$\begin{aligned} X_1(t) = & e^{-\alpha_1 t} X_1(0) + \frac{\alpha_1}{\alpha_2 - \alpha_1} (e^{-\alpha_1 t} - e^{-\alpha_2 t}) X_2(0) \\ & + \int_0^t \alpha_2 \mu \frac{\alpha_1}{\alpha_2 - \alpha_1} (e^{-\alpha_1(t-u)} - e^{-\alpha_2(t-u)}) dt \\ & + \int_0^t \sigma_1 e^{-\alpha_1(t-u)} dW_1(u) + \int_0^t \sigma_2 \frac{\alpha_1}{\alpha_2 - \alpha_1} (e^{-\alpha_1(t-u)} - e^{-\alpha_2(t-u)}) dW_2(u) \end{aligned}$$

hvor første linje viser den deterministiske påvirkningen av utgangspunktet, tilstanden ved null-tidspunktet, linje to viser forventningens påvirkning, og siste linje viser virkningen av begge stokastiske elementer, både det direkte tilhørende topplagsprosessen, og det filtrerte, tilhørende den underliggende prosessen.

Den underliggende prosessen,  $X_2(t)$ , er enklere:

$$X_2(t) = e^{-\alpha_2 t} X_2(0) + \int_0^t \alpha_2 \mu e^{-\alpha_2(t-u)} du + \int_0^t \sigma_2 e^{-\alpha_2(t-u)} dW_2(u)$$

Dette er hva som også brukes som ett-lags-modell.

## 4 Den grunnleggende tankegangen i modelleringen

En rask oppsummering av de grunnleggende trekkene ved en modell for dynamisk og stokastisk interaksjon mellom inntektsulikhet og velferdsstatsgenerøsitet kan være på sin plass: Modellen består av stokastiske differensialligninger, hvor hver av dem har to parametre; en drift-parameter  $\alpha$ , og en diffusjonsparameter  $\sigma$ . Driftparameteren angir hvor raskt størrelsen beveger seg mot det den følger, det være seg en konstant, altså forventningsverdien,  $\mu$ ,

eller en annen prosess, eller en transformering av en prosess. En høy  $\alpha$ -verdien betyr at prosessen bruker lite tid på å innhente prosessen den følger. Diffusjonsparametren multipliserer de standardiserte utfallene fra Wienerprosessen. En høy diffusjonsparameterverdi betyr at prosessen er lite stabil i sin utvikling. Disse konseptene blir beskrevet videre i den økonomiske applikasjonen av modellen.

Hvis vi setter opp et system med  $n$  ligninger, hver med to parametre, kan vi danne  $2^n$  strukturelt forskjellige modelleringer, bare ved hjelp av parameterspesifisering. Med mindre vi velger å la  $dW(t)$ -vektoren være  $n \times 1$ , med uavhengige stokastiske elementer på hver linje, så kan vi få enda flere spesifikasjoner. Vi kan for eksempel la flere linjer få samme stokastiske bidrag. Ikke alle av modellspesifikasjonene er like interessante fra et økonomisk synspunkt. Modelleringen avhenger nødvendigvis av hva man vil ha besvart.

Moene og Barth tar utgangspunkt i Esping-Andersens *Three Worlds of Welfare Capitalism* [Esping-Andersen, 1990], når de inndeler landene i politiske regioner:

- Liberale:
  - Australia, Canada, Irland, New Zealand, Storbritannia, USA.
- Konservative:
  - Finland, Frankrike, Tyskland, Italia, Japan, Sveits
- Sosialdemokratiske:
  - Østerrike, Belgia, Danmark, Nederland, Norge, Sverige.

Ornstein-Uhlenbeck-prosessen har standardnormalfordelt forstyrrelse,  $dW(t)$ . Fordelingen for denne er  $N(0, dt)$ , og bevegelsesområdet er fra  $-\infty$  til  $+\infty$ . For å kunne bruke modellen må det som modelleres ha samme teoretiske bevegelsesgrunn. Datamaterialet er inntektsulikhet og velferdsstatsgenerøsitet for 18 forskjellige land, for forskjellige konsekutive eller ikke-konsekutive år. Inntektsuliketen er definert som niende desil delt på første desil i fordelingen

av timelønn før skatt. Dens teoretisk laveste verdi er 1, som oppnås ved helt like lønninger for alle:  $\frac{d_0}{d_1} = 1$ .

Til modelleringen trengs observasjoner som teoretisk beveger seg på hele tall-linjen. Følgende skala løser begge problemene:  $\ln(\frac{d_0-d_1}{d_1}) = \ln(\frac{d_0}{d_1} - 1)$ .

Det andre datasettet er et mål for velferdsstatsgenerøsitet, hvis teoretiske bevegelsesområde er de positive tallene (evt med null inkludert). Ved å bruke  $\ln$ -transformasjon også her, får vi igjen bevegelsesområde fra  $-\infty$  til  $+\infty$ .

Ved å bruke disse transformerte størrelsene i dette modelleringsrammeverket, hvor forstyrrelsene er normalfordelte, gjør vi indirekte antagelsen at forstyrrelsene for originalstørrelsene er lognormalfordelte. Og det gir mening for størrelser som begynner på null. For en gitt forventningsverdi, så er det begrenset hvor langt ned prosessen kan gå. Inntektsulikheten (minus én) kan definisjonsmessig ikke gå under null, og det samme gjelder velferdstatsgenerøsiteten. Det gir kanskje økonomisk mening å hevde det også er begrenset hvor høyt disse størrelsene kan gå. Men det går heller ikke å sette en absoutt grense for dette. I modelleringsarbeidet tar vi kun hensyn til de teoretiske begrensningene, ikke de praktiske. En normalfordelt variabel vil uansett, på lang sikt, ha 95 % av sine realiseringer innenfor to standardavvik i begge retninger fra forventningen. Vi vil da “aldri”, ved for eksempel simulering, få utfall som plutselig sender inntektsulikheten til det uendelige. Dette teoretiske bevegelsesområdet har altså lite å gjøre med det vi i praksis kan forvente å se.

Datamaterialet går over perioden 1976-2002, og vi har størrelser for 18 land. Tidsseriene er på ingen måte komplette, og for noen land har vi bare fire observasjoner (Sveits).

De 18 landene deles inn i tre kategorier, som hver består av 6 land, slik Moene og Barth gjør det.

## 5 Beskrivelse av den numeriske analysen

Selve C-koden er lagt til et appendix, men en punktvís beskrivelse av optimeringen er verdt å ha med:

1. Innlesing av data
2. Trekke initialparametre tilfeldig, som startparametre for å lete etter optimum
3. Bruke disse initialparametrene til å danne en  $A$ -matrise.
4. Egenvektordekomponere  $A$ -matrisen og bruke dekomponeringen i den teoretiske kovariansmatrisen og forventningsvektoren for observasjonene
5. Basert på parametrene og modellspesifikasjonen beregnes forventningsvektor og kovariansmatrise for alle observasjonene. Ved for eksempel 130 innleste observasjoner av inntekstulikhet og velferdsstatsgenerøsitet har forventningsvektoren dimensjon  $260 \times 1$ , mens kovariansmatrisen har dimensjon  $260 \times 260$ . Hvert element i disse er en funksjon av parametrene og den modellstruktur som er bestemt.
6. Numerisk likelihoodberegning hvor de parameteravhengige forventningsvektoren og kovariansmatrisen sammen med observasjonsvektor gir en likelihoodverdi
7. Optimeringen skjer ved å endre marginalt parameterverdiene og gå i den retning hvor likelihooden øker mest. For hver marginale endring blir punkt 3 repetert, altså ny  $A$ -matrise blir beregnet.
8. Når likelihooden ikke øker lengre ved endringer i parametrene, så tolkes parametrene som optimale og resultatet blir presenter, sammen med antallet runder med marginale endringer.
9. Punkt 1-8 blir repetert et bestemt antall ganger, for eksempel 100, hvor poenget er å satse på at mange av dem finner den globale toppen

## 6 Modelling

### 6.1 Referansemmodell

Den grunnleggende tankegangen i modelleringen ble nevnt tidligere. En optimal beregning av interaksjonsparametrene krever en modell hvor *ingen andre elementer enn nettopp selve interaksjonen* skal forklares av disse parametrene. I så tilfelle må modellstrukturen være slik at andre parametre fanger opp disse andre virkningene. En feilformulert modell vil gjøre at andre virkninger gjerne “presses inn” i interaksjonsparametrene. I en modell for  $k$  land vil jeg dermed sette interaksjonsparametrene lik for alle land, mens hvert land får spesifikke drift- og diffusjonsparametre, samt forventningsverdi, og alt dette for både inntektsulikhet og velferdsstatsgenerøsitet. Dermed vil et vilkårlig antall land kunne brukes til å beregne nettopp  $a_I$  og  $a_G$ . Siden inntektsulikhet og velferdsstatsgenerøsitet gjerne er størrelser som endrer seg over tid, legges det inn en trendfaktor som et felles underliggende lag. Dette vil fange opp de globale faktorer som for eksempel økonomisk vekst, eller et større globalt fokus på nettopp inntektsulikhet og velferdsstatsgenerøsitet. Dermed vil det kunne dannes en langsiktig trend omkring de konstante forventningsverdier, som forklarer at disse vokser (eller synker) over den tidsperioden vi har observasjoner fra. Dette underliggende laget er modellert som de to nederste ligningene i modellen.

$$\begin{aligned}dI_k(t) &= -\alpha_{I_k}(I_k(t) - \mu_{I_k} - a_G(G_k(t) - \mu_{G_k}) - I_u)dt + \sigma_{I_k}dW_{I_k}(t) \\dG_k(t) &= -\alpha_{G_k}(G_k(t) - \mu_{G_k} - a_I(I_k(t) - \mu_{I_k}) - G_u)dt + \sigma_{G_k}dW_{G_k}(t) \\dI_u(t) &= -\alpha_{I_u}I_u(t)dt + \sigma_{I_u}dW_{I_u}(t) \\dG_u(t) &= -\alpha_{G_u}G_u(t)dt + \sigma_{G_u}dW_{G_u}(t)\end{aligned}$$

hvor  $k = 1, 2, \dots, 6$ , slik at vi her har 14 ligninger for 6 land. Ligningene sier i ord sier at inntektsulikheten trekker mot en sum av egen forventning, og den underliggende prosessen for inntektsulikhet, samt generøsitetens avstand fra egen forventning, med  $-a_G$  som proporsjonalitetsfaktor. For generøsiteten er situasjonen symmetrisk. Det underliggende laget, både for inntektsulikhet og

velferdsstatsgenerøsitet, er prosesser uten forventning (altså forventning lik 0). De underliggende prosessene, representert ved de to siste ligningene, er tenkt å forklare langsiktige trender, og å bidra med store stokastiske bidrag. Strukturen er lik for alle land, slik at skillene kun kommer i form av forskjellige parametre. Hvis alle  $\alpha$ -verdier er positive, og det samme for interaksjonsparementene,  $a_I$ , og  $a_G$ , så får vi en modell hvor inntektsulikheten går ned hvis generøsiteten ligger over sin forventning, og hvor generøsiteten går opp hvis inntekstulikheten ligger under sin forventning.

Dette systemet gir følgende  $A$ -matrise,  $m$ -vektor, og  $\Sigma$ -matrise:

$$A = \begin{pmatrix} -\alpha_{I_1} & \alpha_{I_1} a_G & & & & & -\alpha_{I_1} & & \\ \alpha_{G_1} a_I & -\alpha_{G_1} & & & & & & -\alpha_{G_1} & \\ & & -\alpha_{I_2} & \alpha_{I_2} a_G & & & -\alpha_{I_2} & & \\ & & \alpha_{G_2} a_I & -\alpha_{G_2} & & & & -\alpha_{G_2} & \\ & & & & \ddots & \ddots & \vdots & \vdots & \\ & & & & \ddots & \ddots & \vdots & \vdots & \\ & & & & & & -\alpha_{I_6} & \alpha_{I_6} a_G & -\alpha_{I_6} & \\ & & & & & & \alpha_{G_6} a_I & -\alpha_{G_6} & & -\alpha_{G_6} \\ & & & & & & & & -\alpha_{I_u} & \\ & & & & & & & & & \alpha_{G_u} \end{pmatrix}$$



$$m = \begin{pmatrix} \alpha_{I_1}(\mu_{I_1} - a_G \mu_{G_1}) \\ \alpha_{G_1}(\mu_{G_1} - a_I \mu_{I_1}) \\ \alpha_{I_2}(\mu_{I_2} - a_G \mu_{G_2}) \\ \alpha_{G_2}(\mu_{G_2} - a_I \mu_{I_2}) \\ \vdots \\ \vdots \\ \alpha_{I_6}(\mu_{I_6} - a_G \mu_{G_6}) \\ \alpha_{G_6}(\mu_{G_6} - a_I \mu_{I_6}) \\ 0 \\ 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_{I_1} & & & & & \\ & \sigma_{G_1} & & & & \\ & & \sigma_{I_2} & & & \\ & & & \sigma_{G_2} & & \\ & & & & \ddots & \\ & & & & & \ddots & \\ & & & & & & \sigma_{I_6} \\ & & & & & & & \sigma_{G_6} \\ & & & & & & & & \sigma_{I_u} \\ & & & & & & & & & \sigma_{G_u} \end{pmatrix}$$

$$X(t) = \begin{pmatrix} I_1(t) \\ G_1(t) \\ I_2(t) \\ G_2(t) \\ I_3(t) \\ G_3(t) \\ \vdots \\ \vdots \\ I_6(t) \\ G_6(t) \\ I_u(t) \\ G_u(t) \end{pmatrix}$$

Denne referansemodellen har 6 parametre for hvert land, samt to interaksjonsparametre, og fire parametre tilknyttet det underliggende laget. I tillegg forutsettes målefeil. Målefeil knyttet til metoder gir oss 2 ekstra parametre, da én metode er brukt til å bestemme velferdsstatsgenerøsitet, og én metode er brukt for inntektsulikhetsmålinger. Ved bruk av datamateriale for 6 land blir dette dermed en modell med 44 parametre. Det er meningsløst å skulle bestemme så mange parametre, og forenklete modeller følger under. Det er likevel hensiktsmessig å ha denne som referanse, slik at forenklingen lettere kan settes i sammenheng.

I alle de følgende modellene blir det kjørt optimering for hver “velferdsstatsverden” separat. For globale  $a_I$  og  $a_G$  må analysen kjøres på alle land samtidig. Det er fullt mulig, men tar betraktelig lengre tid, og er aktuelt å gjøre hvis de mindre analysene gir relevante svar.

### 6.1.1 Modell 1

Heller enn å ha landsspesifikke  $\alpha$ - og  $\sigma$ -verdier for inntektsulikheter og velferdsstatsgenerøsitet, er disse redusert til fellesparametre, altså  $\alpha_I$ ,  $\alpha_G$ ,  $\sigma_I$  og  $\sigma_G$ . I tillegg blir alle forventningsverdiene fjernet. Dette er en betydelig endring, og det må reflekteres i datamaterialet, som blir “sterilisert” ved at alle landsspe-

sifikke karakteristika blir fjernet. Alle data får fjernet “eget” gjennomsnitt. For eksempel blir gjennomsnittet av inntektsulikhet for et land over observasjonsperioden regnet ut, og alle data fra denne observasjonsperioden får dette fratrasket. For å korrigere for forskjellige lands ulike variasjon, blir alle data skalert ved å dele på eget gjennomsnitt. Dette begrunner begrensningen til to forskjellige  $\sigma$ -parametre.

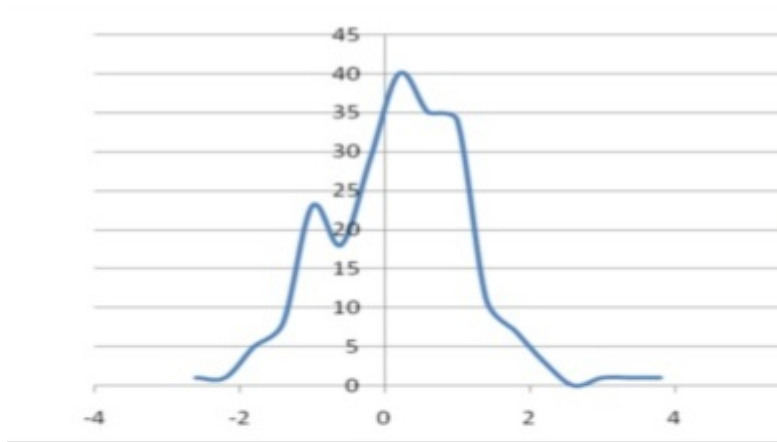
Forventningen for alle prosessene er dermed 0, og det hjelper på den numeriske analysen. Parametrene som skal bestemmes er da:  $\alpha_I$ ,  $\alpha_G$ ,  $\sigma_I$ ,  $\sigma_G$ ,  $\alpha_{I_u}$ ,  $\alpha_{G_u}$ ,  $\sigma_{I_u}$ ,  $\sigma_{G_u}$ ,  $a_I$ ,  $a_G$ ,  $s_1$ ,  $s_2$ , hvor  $s_i$  er standardavviket til selve målingene, og legges på variansen til målingene, hvor I er inntektsulikhet, G er velferdsstatsgenerøsitet, og subscript u betegner det underliggende laget.

Hvorvidt antagelsen om normalfordeling, som i sin natur ikke hører hjemme i noen ting som har med menneskelig *handling* å gjøre, er begrunnet, kan vi se ved å undersøke historgrammer over de transformerte dataene for hver av de tre velferdsstatsverdenene; se figur 1. Mens dette ser ut til være en akseptabel tilnærming for de sosialdemokratiske og de liberale landene, mangler det en del på midten for de konservative landene.

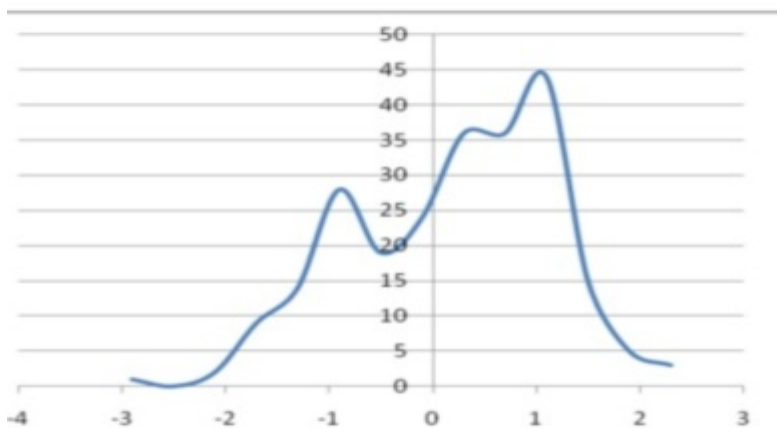
Etter noen hundre optimeringer er konklusjonen for denne modellen at den er flertoppet, og trolig også har mange flate områder i likelihood-overflaten. En analyse kunne konkludere motsatt av forrige, og det var ikke mulig å bestemme et parametersett i dette tolv-dimensjonale parameterrommet. En gjenganger i en del tilfeller var likevel at  $\alpha$ -verdiene for det underliggende laget fikk en lavere verdi enn i topplaget. Dette gir mening. Det underliggende laget skal forklare tendenser som landene har felles, altså regionale trender. En lav  $\alpha$ -verdi vil da kunne fange opp de litt mer langsiktige trendene i datamaterialet (for eksempel en periode med størrelser over gjennomsnittet). Stokastikk i det nederste laget vil dermed kunne henge igjen i modellen lenge.

### 6.1.2 Modell 2

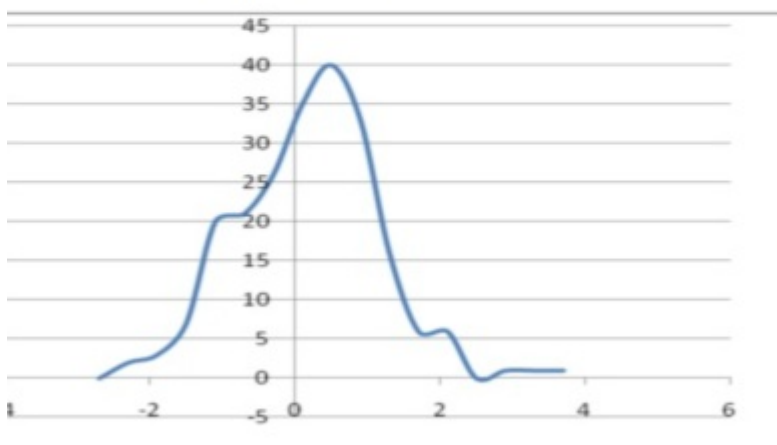
Det underliggende laget skal bidra til å forklare felles tendenser mellom landene. Dermed har topplagsparametrene mindre tyngde og kan bli mer spesi-



liberale land



konserverve land



sosialdemokratiske land

fikke på nettopp dét de er tenkt å forklare. Siste modellstruktur skiller seg fra forrige i at det underliggende laget er fjernet. Dermed finnes det ikke lenger i modellen felleselementer som kan bidra til å forklare observasjonenes kovarians. Hvert land går sin egen vei, uavhengig av nabolandene (nabo i betydningen å tilhøre samme velferdsstatsverden). Likevel er parametersettet krevd å være felles for landene. Med så mange steg unna referansemodellen, skriver jeg opp igjen her  $A$ -matrisen.  $\Sigma$ -matrisen er annenhver  $\sigma_I$  og  $\sigma_G$  på diagonalen.  $m$  er en vektor bestående av null-verdier.

$$A = \begin{pmatrix} -\alpha_I & \alpha_I a_G & & & & \\ \alpha_G a_I & -\alpha_G & & & & \\ & & -\alpha_I & \alpha_I a_G & & \\ & & \alpha_G a_I & -\alpha_G & & \\ & & & & \ddots & \ddots \\ & & & & \ddots & \ddots \\ & & & & & & -\alpha_I & \alpha_I a_G \\ & & & & & & \alpha_G a_I & -\alpha_G \end{pmatrix}$$

Den numeriske analysen skiller seg fra de forrige optimeringene på ett viktig punkt. Den begynner aldri i flate områder. Hvis den begynner i et flatt område, så stopper optimeringen, fordi det ikke er noen “retning” å gå i, hvor likelihooden er høyere. I denne langt enklere modellen er likelihood-flaten i det minste ikke flat, men den er likevel ikke én-toppet. I ett såpass bratt likelihood-område gjør optimeringen gjerne flere hundre bevegelser i parameterrommet før den konkluderer. En viss truktur på de optimale punktene er nå tydelige, men de gir lite mening likevel.

Noen spesifikke resultater er lagt til appendixet.

For interaksjonsparametrene,  $a_I$  og  $a_G$ , blir gjengangerverdier -1, 1 og 0 som resultat. Mulighetsområdet er begrenset til -1, 1, slik at optimeringen når ekstreme verdier i de fleste tilfellene. To forskjellige optimeringer kan gi motsatt resultat av hverandre, for disse parametrene. Det tyder på et

identifikasjonsproblem, og dette vil jeg komme tilbake til.

En forutsetning for stasjonaritet i fordelingen er at egenverdiene er negative. I denne siste modellen, som er den enkleste, er det lett å finne analytisk uttrykk for egenverdiene. Siden hvert land nå er uavhengig av de andre, men hvor de likevel har samme parameterverdi, kan det settes opp en modell for ett enkelt land, og situasjonen er lik for de andre.

Modellen er:

$$A = \begin{pmatrix} -\alpha_I & \alpha_I a_G \\ \alpha_G a_I & -\alpha_G \end{pmatrix}$$

Egenverdiene er:  $\lambda_{1,2} = \frac{-(\alpha_I + \alpha_G) \pm \sqrt{(\alpha_I - \alpha_G)^2 + 4a_I a_G \alpha_I \alpha_G}}{2}$ , og stasjonaritet krever at alle egenverdiene er negative. I det ene tilfellet, hvor kvadratroten trekkes fra, er det åpenbart innfridd, mens det andre tilfellet stiller krav: Vi må ha:  $|a_I \times a_G| < 1$ . I analysen er det innfridd.

En optimering er også gjort uten å begrense bevegelsesrommet for interaksjonsparametrene til  $(-1, 1)$ . Denne er først relevant å diskutere etter å ha presentert modell 3.

En annen endring ble også forsøkt. Heller enn å fjerne gjennomsnittet, ble en lineær trend fjernet. Dette skal bedre reflektere det at et underliggende lag er fjernet. Når trenden er fjernet, finnes det ikke mer enn bevegelse omkring (lineær) trend som kan forklares. Men heller ikke denne transformasjonen av datamaterialet var til hjelp. Optimeringen stopper ganske raskt, og interaksjonsparametrene estimeres gjerne til -1, 1, eller 0, og dette gir ikke mening, spesielt med tanke på at de er begrenset til å befinne seg i intervallet  $(-1, 1)$ .

## 6.2 Modell 3

Modell 2 er forenklet ved at interaksjonen blir fjernet. Dermed modelleres inntektsulikhet og velferdsstatsgenerøsitet som uavhengige av hverandre. Multimodalitetsproblemet er kraftig redusert, og de fleste optimeringene kommer til samme konklusjon. Disse er lagt til appendiks avsnitt 8.7. En naturlig konsekvens av det er at spørsmålene en kan få besvart ikke lenger er de samme som med interaksjon. Spørsmålene som nå lar seg besvare handler direkte om tidsaspektet ved utviklingen av inntektsulikheten og velferdsstatsgen-

erøstiteten. Et hensiktsmessig konsept er i den sammenhengen “karakteristisk tid”. Det er tiden det tar før korrelasjonen synker til  $e^{-1}$ . Det kan til en viss grad regnes som tiden det tar før fortiden “er glemt”. Det mest interessante er likevel å sammenligne karakteristisk tid for forskjellige variable for forskjellige regioner, for å danne seg et bilde over “hastigheten”, eller “tregheten” i systemene.

Etter at datamaterialet er detrendet og skalert (ved å dele på eget standardavvik), så finnes det ikke mye mer enn fluktuasjoner omkring trend igjen å utforske. En modellutvidelse ble forsøkt, hvor et underliggende lag ble lagt til, men fremdeles ingen interaksjon. Det hadde ikke signifikant bedre forklaringskraft enn den enkle modellen, og konklusjonene her hentes dermed fra den enkleste varianten. De to modellvariantene har for samme datamateriale et *likelihood ratio* på omtrent én. Dermed bør den enklere modellen velges. At et ekstra underliggende lag ikke bidro til forklaring, betyr at det ikke fantes noe mønster med lange trender, utover den lineære trenden som er fjernet. Det betyr at veksten, eller reduksjonen, i inntektsulikhet og velferdsstatsgenerøsitet, er av en stabil art, som har relativt korte trender omkring trend, altså korte nok til å forklares av en én-dimensjonal OU-prosess.

	<i>Liberale land</i>	<i>Konservative land</i>	<i>Sosialdemokratiske land</i>
$K_{\alpha_I}$	1.71	2.77	1.35
$K_{\alpha_G}$	1.12	4.45	1.42

hvor  $K_{\alpha_I}$  er karakteristisk tid for inntektsulikheten.

Det tydelige mønsteret her er at de konservative landene har større “treghet” i begge størrelser. For de liberale landene og de sosialdemokratiske landene er stabiliteten større. Fluktuasjoner omkring trend for disse størrelsene er kortvarige. Det vi ser på her er altså logaritmisk transformerte størrelser, så det vi ser på i bunn og grunn er fluktuasjoner i veksten i seg selv. Her viser altså de konservative landene å være mer sensitive for ytre forstyrrelser enn hva gjelder de to andre velferdsstatsverdenene; de bruker altså lengre tid på “å komme” seg etter et ytre sjokk.

Hva kan så være årsaken til den manglende troverdige identifikasjonen av

interaksjonsparametrene? Det umiddelbare svaret er multimodalitet i likelihoodflaten. Men hva er så årsaken til denne multimodaliteten? Altså at der finnes parametervektorer som gir en bedre likelihood enn hva parametervektorer “i nærheten” gir, slik at der ikke finnes noen jevnt stigende “sti” til toppen? En rask sammenligning av kovariansstrukturen, som jo er nettopp det som ligger til grunn for likelihood-beregningen, kan gi et hint: For ett-lags-modellen uten interaksjon, vil kovariansen ha følgende struktur, altså for en og samme prosess, mellom tidspunktene  $v$  og  $t$ :

$$\text{Cov}(X(v), X(t)) = E \left[ e^{-\alpha(t+v)} \sigma^2 \left[ \int_0^v e^{-as} dW(s) \right]^2 \right]$$

mens kovariansen for samme tidspunkter, for en prosess som interagerer med en annen prosess, er følgende, med  $\alpha$  og  $\beta$  som to drift-parametre, og  $a$  og  $b$  som interaksjonsparametre:

$$\begin{aligned} \text{Cov}(X(v), X(t)) = & E \left[ \sigma_1 \int_0^v A e^{\lambda_1(t-u)} + B e^{\lambda_2(t-u)} dW_1(u) + \right. \\ & \left. \sigma_2 \int_0^v A e^{\lambda_1(t-u)} + B e^{\lambda_2(t-u)} dW_2(u) \right] \times \\ & \left[ \sigma_1 \int_0^v A e^{\lambda_1(v-u)} + B e^{\lambda_2(v-u)} dW_1(u) + \right. \\ & \left. \sigma_2 \int_0^v A e^{\lambda_1(v-u)} + B e^{\lambda_2(v-u)} dW_2(u) \right] \end{aligned}$$

hvor

$$\begin{aligned} A &= \frac{a\alpha(\alpha + \lambda_1)}{(a\alpha b\beta - (\alpha + \lambda_1)(\beta + \lambda_2))} \\ B &= \frac{-(\alpha + \lambda_1)(\beta + \lambda_2)}{(a\alpha b\beta - (\alpha + \lambda_1)(\beta + \lambda_2))} \\ E &= \frac{b\beta}{\alpha + \lambda_1} \\ H &= \frac{a\alpha}{\beta + \lambda_2} \\ \lambda_{1,2} &= \frac{-(\alpha + \beta) \pm \sqrt{(\alpha - \beta)^2 + 4a\alpha b\beta}}{2} \end{aligned}$$



Å håpe på en smidig funksjon med interaksjon, slik vi gjerne ønsker i vårt fag, er tilsynelatende å håpe forgjeves.

## 7 Konklusjon

Oppgavens tittel er: *En undersøkelse av hvorvidt systemer av stokastiske differensialligninger egner seg for å undersøke utviklingen av, og interaksjon mellom, inntektsulikhet og velferdsstatsgenerøsitet.* En foreløbig konklusjon blir dermed at de egner seg for undersøkelse av utviklingen av, men ikke av interaksjonen mellom. Hvorvidt dette skriver seg fra feilvalgt optimeringsmetode eller manglende identifikasjon av interaksjonsparametrene er usikkert. Det som er sikkert er at metoden bruker en kovariansmatrise til å beregne parameterverdiene. Hvis den matematiske struktur på denne matrisen er slik at den inneholder frihetsgrader i parametrene, for en optimal matrise, altså med fastlåste verdier, så har vi manglende identifikasjone. Jeg regnet ut denne matrisen, og en begynnelse på dette er gjort ovenfor, og fant at interaksjonsparametrene i de fleste tilfeller forekom i matrisen som et direkte produkt. Et produkt har alltid frihetsgrader. I en lang sum forekom den som produkt mange ganger, men i ett av tilfellene befant den ene parametren seg alene. Med to ligninger for to parametre bør de være identifiserte. Men vi kan ha semi-identifikasjon, og den kan være vanskelig å finne i praksis. Iallefall gikk det ikke ved hjelp av denne måten. De egner seg for studie av utvikling generelt, og da også utvikling i velferdsstatsgenerøsitet og inntektsulikhet. Noen numeriske konklusjoner foreligger, som oppsummert ovenfor, og i sin helhet i appendix (avsnitt 8.7).

## 8 Appendix

### 8.1 Førsteordens lineære differensialligninger

#### 8.1.1 Deterministisk

En enkel variant av en førsteordens lineær differensialligning er følgende:

$$dX(t) = -aX(t)dt, \quad a > 0, \quad X(0) \neq 0, \quad \text{gitt}$$

Hvis denne ligningen beskriver en prosess, kommer vi umiddelbart til noen enkle konklusjoner: Veksthastigheten er proporsjonal med prosessens avstand fra null (i x-aksens retning), og hvis utgangspunktet er positivt er prosessen synkende, og motsatt ellers.

En enkel utvidelse av modellen får vi ved å legge til en konstant på følgende måte:

$$dX(t) = -a(X(t) - b)dt, \quad b \text{ gitt konstant}$$

Denne utvidelsen skiller seg fra den første varianten ved det at referanseverdien ikke lenger er 0, men  $b$ . Ved å sette  $b = 0$  gjenfinner vi det første eksemplet. Veksthastigheten her er proporsjonal med avstanden til  $b$ .

En ytterligere utvidelse får vi ved å skifte ut konstanten  $b$  med en funksjon,  $f(t)$ :

$$dX(t) = -a(X(t) - f(t))dt, \quad f \text{ en gitt funksjon}$$

Ved samme tankegang som tidligere, ser vi at prosessens veksthastighet på et gitt tidspunkt, er proporsjonal med avstanden til  $f$ -funksjonen, og videre at prosessen synker når avstanden er positiv, og stiger ellers. Prosessen vil dermed “følge” funksjonen  $f$  på ethvert tidspunkt. Dette kan vi kalle en følgeprosess.

For å finne den funksjonen som løser differensialligningen, multipliserer vi hvert ledd med den integrerende faktor  $e^{at}$ , flytter de elementene av ligningen som inneholder  $X$  over på venstre side, og konstaterer at vi der da finner den

deriverte av uttrykket  $X(t)e^{at}$ . Trinnvis ser det slik ut:

$$\begin{aligned}d(X(t))e^{at} &= -aX(t)e^{at}dt + af(t)e^{at}dt \\d(X(t))e^{at} + aX(t)e^{at}dt &= af(t)e^{at}dt \\d(X(t)e^{at}) &= af(t)e^{at}dt \\X(t)e^{at} &= a \int f(t)e^{at}dt + C \\X(t) &= e^{-at}a \int f(t)e^{at}dt + e^{-at}C\end{aligned}$$

hvor eventuelle gitte initialbetingelser bestemmer konstanten  $C$ .

### 8.1.2 Stokastikk

De foregående eksemplene viste deterministiske prosesser, hvor utgangspunktet og tiden var nok til å beskrive prosessen i sin helhet. Hvis den presisjonsgraden i liten grad samsvarer med et fenomen vi ser for oss å modellere, så kan vi legge til forstyrrelser, eller usikkerhetsselementer. Ved å legge til en standardisert Wiener-prosess,  $W(t)$ , får vi innfridd det ønsket. Det er for øyeblikket mest hensiktsmessig med en intuitiv forklaring av hva dette fenomenet er.

## 8.2 Momenter for en én-dimensjonal Ornstein-Uhlenbeck-prosess

- Integrering av en én-dimensjonal Ornstein-Uhlenbeck-prosess:

$$dX(t) = -a(X(t) - \mu)dt + \sigma dW(t)$$

$$dX(t) + aX(t)dt = a\mu dt + \sigma dW(t)$$

$$dX(t)e^{at} + aX(t)e^{at}dt = a\mu e^{at}dt + \sigma e^{at}dW(t)$$

$$d(X(t)e^{at}) = a\mu e^{at}dt + \sigma e^{at}dW(t)$$

$$X(t)e^{at} = C + \int_0^t a\mu e^{au}du + \int_0^t \sigma e^{au}dW(u)$$

$$X(t) = e^{-at}C + e^{-at}a\mu \int_0^t e^{au}du + e^{-at}\sigma \int_0^t e^{au}dW(u)$$

$$X(t) = e^{-at}C + e^{-at}a\mu \frac{1}{a}e^{at} + e^{-at}\sigma \int e^{at}dW(t)$$

$$X(t) = e^{-at}C + \mu + e^{-at}\sigma \int_0^t e^{au}dW(u)$$

Hvis vi antar at sannsynlighetsfordelingen for observasjonen på 0-tidspunktet er lik sannsynlighetsfordelingen på de andre tidspunktene, så har vi en stasjonær sannsynlighetsfordeling. Det er en rimelig antagelse å gjøre, at det tidspunktet vi (gjærne tilfeldigvis) begynner våre observasjoner, ikke har direkte betydning for hvordan fordelingen skal være. Ved stasjonær fordeling bestemmer vi følgende:

$$X(t) = e^{-at}C + \mu + e^{-at}\sigma \int_0^t e^{as}dW(s)$$

$$X(0) = e^{-a0}C + \mu + e^{-a0}\sigma \int_0^0 e^{at}dW(t)$$

For bestemmelse av konstanten  $C$  bruker vi følgende:

$$X(0) = C + \mu \rightarrow C = X(0) - \mu$$

Som gir oss følgende prosessbeskrivelse:

$$X(t) = e^{-at}(X(0) - \mu) + \mu + e^{-at} \int_0^t \sigma e^{as} dW(s)$$

Som vi igjen benytter for å bestemme prosessens forventning:

$$E(X(t)) = E(e^{-at}(X(0) - \mu) + \mu)$$

hvor vi kun kan ha  $E(X(0)) = E(X(t))$  hvis begge tar verdien  $\mu$ .

Proessen er altså stasjonær med forventning  $\mu$ .

- Varians for en én-dimensjonal Ornstein-Uhlenbeck-prosess

$$\begin{aligned} \text{Var}(X(t)) &= E[(X(t) - \mu)^2] \\ &= E[e^{-at}(X(0) - \mu) + \sigma e^{-at} \int_0^t e^{as} dW(s)]^2 \\ &= E[e^{-at}(X(0) - \mu)]^2 + 2E[e^{-at}(X(0) - \mu)\sigma e^{-at} \int_0^t e^{as} dW(s)] + \\ &\quad E[\sigma^2 e^{-2at} \int_0^t e^{as} dW(s)]^2 \\ &= e^{-2at} E[(X(0) - \mu)^2] + \sigma^2 e^{-2at} \int_0^t e^{2as} ds \\ &= e^{-2at} \text{Var}(X(0)) + \frac{\sigma^2}{2a} e^{-2at} (e^{2at} - 1) \\ &= e^{-2at} \text{Var}(X(0)) + \frac{\sigma^2}{2a} (1 - e^{-2at}) \\ &= \frac{\sigma^2}{2a} + e^{-2at} (\text{Var}(X(0)) - \frac{\sigma^2}{2a}) \end{aligned}$$

hvor vi ser at den stasjonære fordelings beskrivelse ved:  $\text{Var}(X(t)) = \text{Var}(X(0)) = \frac{\sigma^2}{2a}$ . Det en kan merke seg at forventningen til kvadratet av et stokastisk integral, er lik det deterministiske integralet av kvadratet av integranden. Dette er en direkte konsekvens av konstruksjonen av Wiener-proessen, og nettopp denne egenskapen kalles Itô-isometrien. For beskrivelse av denne henviser jeg til appendikset.

Kortfattet sammendrag er altså følgende, under stasjonaritetsforutset-

ning:  $E(X(t)) = \mu$  og  $Var(X(t)) = \frac{\sigma^2}{2a}$ .

- Stasjonær kovarians for en én-dimensjonal Ornstein-Uhlenbeck-prosess:

$$\begin{aligned}
Cov(X(t_1), X(t_2)) &= E[(X(t_1) - \mu)(X(t_2) - \mu)] \\
&= E[(e^{-at_1}(X(0) - \mu) + e^{-at_1} \int_0^{t_1} \sigma e^{-as} dW(s))(e^{-at_2}(X(0) - \mu) + \sigma e^{-at_2} \int_0^{t_2} \sigma e^{-as} dW(s))] \\
&= E[e^{-a(t_1+t_2)}(X(0) - \mu)^2] + E[e^{-a(t_1+t_2)} \sigma^2 \left( \int_0^{t_1} e^{-as} dW(s) \right)^2] \\
&= e^{-a(t_1+t_2)} \frac{\sigma^2}{2a} + e^{-a(t_1+t_2)} \sigma^2 \int_0^{t_1} e^{2as} ds \\
&= e^{-a(t_1+t_2)} \frac{\sigma^2}{2a} + e^{-a(t_1+t_2)} \frac{\sigma^2}{2a} (e^{2t_1} - 1) \\
&= e^{-a(t_2-t_1)} \frac{\sigma^2}{2a}
\end{aligned}$$

Slik at korrelasjonen blir:

$$Corr(X(t_1), X(t_2)) = e^{-a(t_2-t_1)}$$

gitt  $t_2 > t_1$ .

### 8.3 Standardisert Wiener-prosess

Den standardiserte Wiener-prosessen er én-dimensjonal og er uten “moment”. “Moment” må her forstås som “tyngde” eller “tendens” i deres intuitive forståelse. Matematisk betyr det at  $W(t)$  ikke har noen derivert. Det er altså ikke mulig å forutse hvor prosessen vil gå i fremtiden, basert på hva en vet på et gitt tidspunkt. Den er en random walk. Prosessen vil i løpet av et tidsrom enten gå opp eller ned. Hvis vi lar både tidsrommet og stegstørrelsen gå mot null, og lar tiden gå, oppnår vi den kontinuerlig fluktuerende prosessen vi kaller Wiener-prosess (eller Brownsk bevegelse). En enkel forklaring på ikke-deriverbarheten får vi ved å bruke definisjonen på den deriverte. Ved bevisst feilnotasjon ser det slik ut:

$$\frac{d}{dt}(W(t)) = \lim_{dt \rightarrow 0} \frac{W(t+dt) - W(t)}{dt} = \pm\infty$$

fordi det konstruksjonsmessig er altså slik at i grensen hvor  $dt \rightarrow 0$ , så er er stegstørrelsen uendelig stor, relativt til tidsperioden, selv om de begge går mot null.

Prosessene er normalfordelt med to viktige egenskaper:

- $E[W(s+t) - W(s)] = 0$
- $Var[W(s+t) - W(s)] = t$

Altså har vi  $[W(t+s) - W(s)] \sim N(0, t)$

Motivasjonen for å introdusere et stokastisk element med nettopp disse egenskapene, må forstås på bakgrunn av sentralgrenseteoremet.

Den enkleste versjonen er slik [fra Matematisk formelsamling for økonomer]: La  $\{X_i\}$  være en følge av uavhengige og identisk fordelte stokastiske variabler med endelig forventning  $E[X_i] = \mu$  og endelig varians  $Var[X_i] = \sigma^2$ . La  $S_n = X_1 + X_2 + \dots + X_n$ . Da vil fordelingen av  $\frac{S_n - n\mu}{\sigma\sqrt{n}}$  konvergere mot standardnormalfordelingen når  $n \rightarrow \infty$ , dvs:

$$P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}}\right\} \leq a \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx$$

Når en setter opp en modell for et fenomen, så er det rimelig å anta at det deterministiske elementet i modellen skal modellere det forventede utfallet av prosessen, som i definisjonen gitt ovenfor tilsvarer  $n\mu$ . Det resterende, som da er  $S_n - n\mu$  kan modelleres som det stokastiske elementet. Vi har  $[S_n - n\mu] \sim N(0, \sigma\sqrt{n})$ , mens  $\sigma_2 W(t) \sim N(0, \sigma_2)$ . For  $n$  i praksis en endelig størrelse, vil denne modelleringen gi mening.

Altså, hvis en prosess kan tenkes å være påvirket av mange faktorer, hvor noen faktorer “trekker” funksjonen ned, mens andre “løfter” den opp, samt at disse faktorene er uavhengige av hverandre, så vil en normalfordelt forstyrrelse være en fornuftig modellering.

Ved å legge til dette elementet oppnår vi følgende ligning:

$$dX(t) = -a(X(t) - b)dt + \sigma dW(t)$$

En prosess beskrevet på denne måten, med både “drift” mot et stabilt punkt,  $b$ , og kontinuerlige forstyrrelser i banen, kalles en Ornstein-Uhlenbeck-prosess, og brukes for å modellere en prosess som fluktuerer omkring et likevektspunkt. Vi kaller  $a$  driftparameteren, og  $\sigma$  diffusjonsparameteren.

Den naturlige utvidelsen fra dette utgangspunktet er å erstatte likevektspunktet  $b$  med en funksjon, på samme måte som over:

$$dX(t) = -a(X(t) - f(t))dt + \sigma dW(t)$$

Disse to sistnevnte ligningene er hva som danner byggesteinene for modelleringen i denne oppgaven. Ved å sette dem sammen i spesifikke systemer kan man modellere den interaksjonen mellom størrelser en ser for seg representere enkle aspekter ved virkeligheten

Det matematiske fundamentet for den standardiserte Wiener-prosessen:

Avskrift fra Seierstad: Stochastic Control in Discrete and Continuous Time, side 185:

Consider a stochastic process in the form of a “symmetric, random walk” of the following type. Assume that at times  $h, 2h, 3h, \dots$  we take a step of size  $c$  up or down with probability  $\frac{1}{2}$ . More precisely, suppose there are independent random variables  $Y_t$ , for which  $Pr[Y_t = c] = Pr[Y_t = -c] = \frac{1}{2}$ , for  $t = h, 2h, 3h, \dots$ . Let  $X_t = Y_h + \dots + Y_{nh}$ , if  $t = nh$  for  $n$  a natural number,  $X_0 = 0$ . Because  $E[Y_t] = 0$  and  $Var[Y_t] = E[Y_t^2] = c^2$ , we get  $E[X_t] = 0$  and  $Var[X_t] = c^2(\frac{t}{h})$  where  $\frac{t}{h} = n$ , the number of terms in the sum defining  $X_t$ . In the interval  $[ih, (i+1)h]$ , define  $X_t = X_{ih}$ . For  $t$  in this interval  $E[X_t] = E[X_{ih}]$ ,  $Var[X_t] = Var[X_{ih}]$ . We shall now let  $h$  and  $c$  tend to 0, and when the convergence of  $h$  and  $c$  is suitably coordinated, the resulting limit  $B_t := \lim_{h \rightarrow 0} X_t$  is a “sensible,” nontrivial process defined for all  $t$ . Let us try  $c = ah^\alpha$ ,  $a > 0$ ,  $\alpha > 0$ ,  $a$  and  $\alpha$  fixed. Evidently,  $Var[X_t]$  vanishes when  $h \rightarrow 0$ , if  $\alpha > \frac{1}{2}$ . If the convergence of  $c$  is too slow, i.e.,  $0 < \alpha < \frac{1}{2}$ , we get infinite variance. However, the choices  $\alpha = \frac{1}{2}$ ,  $c = h^{\frac{1}{2}}$ , and (for simplicity)



$a = 1$ , make  $E[X_t] = 0$  and  $Var[X_t] = t$ . Both equalities hold even in the limit, i.e., both equalities hold for  $\lim_{h \rightarrow 0} X_t =: B_t$ . The so-called central limit theorem in probability theory says that a sum of many small, independent random variables of the type  $Y_k$  is approximately normally distributed, and in the limit exactly so, in fact,  $B_t$  is  $N(0, t)$ , i.e., normally distributed with mean 0 and variance  $t$ . The limit process  $B_t$  is called *Brownian motion* or the *standardized Wiener process*.

## 8.4 The Ito formula

Avskrift fra Seierstad: p196. Consider a stochastic process (sometimes called Ito process) of the form:

$$X_t = X_0 + \int_0^t u(s)ds + \int_0^t v(s)dB_s,$$

where  $u(s)$  and  $v(s)$  are adapted real-valued functions belonging to  $N^*$  and  $t \in J := [0, T]$ . Here  $X_0$  is a fixed number. When a process  $X_t$  is given in this manner, an informal manner of writing it is on differential form,

$$dX_t = u(t)dt + v(t)dB_t, \quad X_0 = x_0, \text{ given}$$

Let  $Y_t := g(t, X_t)$ , where  $g(t, x)$  has continous second derivatives ( $g : R^2 \rightarrow R$ ). Then *Ito's formula* is:

$$Y_t := g(0, X_0) + \int_0^t v(s) \left[ \frac{\partial g(s, X_s)}{\partial x} \right] dB_s + \int_0^t \left\{ \frac{\partial g(s, X_s)}{\partial s} + u(s) \frac{\partial g(s, X_s)}{\partial x} + \frac{1}{2} v(s)^2 \left[ \frac{\partial^2 g(s, X_s)}{\partial x^2} \right] \right\} ds$$

## 8.5 Maximum likelihood

[Avskrift fra Rice: side 267:] Suppose that random vaiables  $X_1, X_2, \dots, X_n$  have a joint density or frequency function  $f(x_1, x_2, \dots, x_n | \theta)$ . Given observed values  $X_i = x_i$ , where  $i = 1, 2, \dots, n$ , the likelihood of  $\theta$  as a function of

$x_1, x_2, \dots, x_n$  is defines as

$$lik(\theta) = f(x_1, x_2, \dots, x_n | \theta)$$

Note that we consider the joint density as a function of  $\theta$  rather as a function of the  $x_i$ . If the distribution is discrete, so that  $f$  is a frequency function, the likelihood function gives the probability of observing the given data as a function of the parameter  $\theta$ . The **maximum likelihood estimate (mle)** of  $\theta$  is that value of  $\theta$  that maximizes the likelihood - that is, makes the observed data “most probable” or “most likely.”.

If the  $X_i$  are assumed to be i.i.d., their joint density is the product of the marginal densities, and the likelihood is

$$lik(\theta) = \prod_{i=1}^n f(X_i | \theta)$$

Rather than maximizing the likelihood itself, it is usually easier to maximize its natural logarithm (which is equivalent since the logarithm is a monotonic function). For an i.i.d. sample, the **log likelihood** is

$$l(\theta) = \sum_{i=1}^n \log[f(X_i | \theta)]$$

## 8.6 Optimeringsutskrifter for modell 2

Her presenteres noen typiske resultater fra optimeringen i modell 2, hvor parameterverdiene blir presentert, sammen med likelihood-verdien. Først kommer verdiene som er trukket som utgangspunkt, sammen med likelihooden, samt antall iterasjoner som er gjort i optimeringen. Deretter kommer slutresultatet. Optimeringen er kjørt på sosialdemokratiske land:

Før:

-loglik: 329.674108

s1=0.374959 s2=0.772644 alphaI=0.491760 alphaG=1.699165

sigmaI=1.701500 sigmaG=0.155359 aI=0.272797 aG=-0.599785 -loglik=329.674108

iterations:697

Etter:

-loglik: 203.515234  
s1=0.001171 s2=0.141256 alphaI=0.006556 alphaG=0.174115  
sigmaI=0.571840 sigmaG=0.002635 aI=-0.999958 aG=0.997641 -loglik=203.515234

Før:

-loglik: 280.547568  
s1=0.703682 s2=0.378311 alphaI=0.382672 alphaG=0.102923  
sigmaI=0.345248 sigmaG=0.318710 aI=-0.660348 aG=0.142878 -loglik=280.547568  
iterations:338

Etter:

-loglik: 203.517581  
s1=0.001303 s2=0.141345 alphaI=0.006528 alphaG=0.174209  
sigmaI=0.571767 sigmaG=0.004036 aI=-0.999948 aG=0.997091 -loglik=203.517581

Før:

-loglik: 311.446945  
s1=0.827459 s2=0.619977 alphaI=0.154452 alphaG=0.041272  
sigmaI=1.663423 sigmaG=0.296746 aI=-0.125437 aG=-0.665598 -loglik=311.446945  
iterations:234

Etter:

-loglik: 180.128491  
s1=0.000624 s2=0.000912 alphaI=0.137536 alphaG=0.003822  
sigmaI=0.001615 sigmaG=0.514383 aI=0.997974 aG=-0.999976 -loglik=180.128491

## 8.7 Resultater for ett og to lag, med og uten interaksjon

Modell A og B her svarer til to-lagsmodellen som er utregnet i avsnitt 3.5, side 18.

Modellbeskrivelse:

A: 2 lag uten interaksjon - gjennomsnittet er fjernet

B: 2 lag uten interaksjon - detrendet data

C: 1 lag med interaksjon - detrendet data

D: 1 lag uten interaksjon - detrendet data

	A	B	C	D
Liberale land				
Minus log likelihood	216,19	294,68	265,66	294,61
$\alpha_I$	0,23	0,57	0,32	0,58
Karakteristisk tid	4,33	1,76	3,74	1,71
$\alpha_G$	0,74	0,86	0,02	0,89
Karakteristisk tid	1,36	1,16	12,35	1,12
$\sigma_I$	0,17	1,03	0,33	1,05
$\sigma_G$	0,28	1,27	0,68	1,30
$a_{IU}$	1,00	0,14		
$a_{GU}$	1,41	0,14		
$a_I$			-0,94	
$a_G$			1,80	
$\alpha_{IU}$	0,22	0,53		
Karakteristisk tid	4,52	1,90		
$\alpha_{GU}$	0,60	0,81		
Karakteristisk tid	1,65	1,23		
$\sigma_{IU}$				
$\sigma_{GU}$				
$s_1$	0,28	0,11	0,19	0,02
$s_2$	0,39	0,14	0,30	0,02
Konservative land				
Minus log likelihood	177,06	254,17	232,90	257,66
$\alpha_I$	0,26	0,58	0,10	0,36
Karakteristisk tid	3,90	1,71	6,66	2,77
$\alpha_G$	0,29	0,60	0,54	0,22
Karakteristisk tid	3,50	1,66	2,42	4,45
$\sigma_I$	0,09	0,14	0,59	0,79
$\sigma_G$	0,26	0,45	0,05	0,72
$a_{IU}$	1,13	1,27		
$a_{GU}$	1,19	1,38		

$a_I$			-1,45	
$a_G$			2,82	
$\alpha_{I_U}$	0,25	0,56		
Karakteristisk tid	3,94	1,77		
$\alpha_{G_U}$	0,26	0,55		
Karakteristisk tid	3,81	1,83		
$\sigma_{I_U}$				
$\sigma_{G_U}$				
s 1	0,34	0,50	0,11	0,31
s 2	0,17	0,20	0,03	0,00

Sosialdemokratiske land

Minus log likelihood	240,25	284,70	231,99	284,69
$\alpha_I$	0,58	0,74	0,10	0,74
Karakteristisk tid	1,73	1,35	10,40	1,35
$\alpha_G$	0,46	0,71	0,03	0,70
Karakteristisk tid	2,18	1,41	28,15	1,42
$\sigma_I$	0,17	1,22	0,16	1,22
$\sigma_G$	0,31	1,17	0,03	1,17
$a_{I_U}$	1,66	0,09		
$a_{G_U}$	1,09	0,07		
$a_I$			0,05	
$a_G$			-3,20	
$\alpha_{I_U}$	0,57	0,74		
Karakteristisk tid	1,75	1,35		
$\alpha_{G_U}$	0,45	0,57		
Karakteristisk tid	2,21	1,75		
$\sigma_{I_U}$				
$\sigma_{G_U}$				
$s_1$	0,32	0,06	0,05	0,02
$s_2$	0,48	0,16	0,29	0,16

## 8.8 C-koden

C-koden følger her presist. Det er en modifikasjon av Trond Reitan sin kode for optimering i en anvendelsen innen biologi. Modellspesifikasjonen er endret, innlesningsrutinen er endret, og selve optimeringsrutinene er endret. De kalkyler som gjøres er ellers beholdt like. Koden er kommentert, og kommentarene finnes etter `//`-tegn. Koden som her følger er den som er brukt i den enkleste modellen, altså ett lag uten interaksjon. Den er kun med for kompletthets skyld, og ikke ment å leses *per se*.

[Reitan, 2010] [Sydsaeter, 2006a] [Moene, 2009]

```
#include <stdio.h>
#include <stdlib.h>
#include <cmath>
#include <time.h>
#include <hydrasub/hydrabase/newton.H>
#include <hydrasub/hydrabase/divfunc.H>
#include <hydrasub/hydrabase/linalg.H>

//Structure for measurements: Political region, country number, country name, year, ge
struct measurements
{
    int reg, ctrnr;
    char ctr[10];
    double year, gen, ineq;
};

// USAGE:
// Printed when the user gives the wrong number of inputs:
void usage(void)
{
    printf("Usage: multi_optim <in file> <#starts>\n");
    exit(0);
}
```

```

// get_measurements:
// Reads the given file into an array of measurements that
// are returned. Also returned is the number of measurements (len)

measurements *get_measurements(char *infile, int *len)
{
    //buffer values for the measurements
    double year, gen, ineq;
    int reg, ctrnr;
    char ctr[10];
    FILE *f=fopen(infile, "r");
    int i,lines=0;

    // check if the file could be opened:
    if(!f)
    {
        printf("Failed to open file %s!\n", infile);
        exit(0);
    }

    // find the number of lines with sensible content:
    while(!feof(f))
    {
        if(fscanf(f,"%d %s %lf %lf %lf %d", &reg, ctr, &year, &gen,&ineq, &ctrnr)==6)
        lines++;
    }
    fclose(f);

    // check if the number of lines was zero:
    if(lines==0)
    {
        printf("No lines read. Illegal format?\n");
        exit(0);
    }
}

```

```

// set the output length:
*len=lines;

// make the output array:
measurements *ret=new measurements[lines];

// read the file content:
f=fopen(infile, "r");
i=0;
while(!feof(f))
{
    if(fscanf(f,"%d %s %lf %lf %lf %d", &(ret[i].reg), ret[i].ctr, &(ret[i].year),
&(ret[i].gen),&(ret[i].ineq), &(ret[i].ctrnr))==6)
{ //prints the input file to the screen
    printf("%d\t %s\t %lf\t %lf\t %lf\t %d\n",
ret[i].reg, ret[i].ctr, ret[i].year,
ret[i].gen, ret[i].ineq, ret[i].ctrnr);
    i++;
}
}

return ret;
}

// global variables for the measurements:
measurements *meas;
int meas_len;

double **get_eigenvector_matrix(double **A, int size, double **lambda)
{
    // Find the eigenvectors and eigenvalues:
    Matrix Amat(A,size,size);
    cmatrix Acmat(Amat);
    complex *lambda_complex;
    cmatrix Vinv_cmat=Acmat.eigenvectors_as_matrix(&lambda_complex);

```



```

int i,j;

// check if there are complex eigenvalues:
for(i=0;i<size;i++)
    if(ABSVAL((lambda_complex[i].Im()))>0.001)
    {
cout << i << "th eigenvalue has imaginary component, lambda" << i << "=( " <<
    lambda_complex[i].Re() << ", " << lambda_complex[i].Im() << ")" << endl;
return NULL;
    }

// fill out the eigenvalue vector:
*lambda=new double[size];
for(i=0;i<size;i++)
    (*lambda)[i]=lambda_complex[i].Re();

// check if there are complex eigenvector values:
for(i=0;i<size;i++)
    for(j=0;j<size;j++)
        if(ABSVAL((Vinv_cmat.get_element(i,j).Im()))>0.001)
{
cout << i << ", " << j << "th eigenvector component has imaginary component, "
    "Vinv_" << i << ", " << j << "=( " <<
    Vinv_cmat.get_element(i,j).Re() << ", " <<
    Vinv_cmat.get_element(i,j).Im() << ")" << endl;
return NULL;
}

// Make the eigenvector matrix:
double **Vinv=new double*[size];
// 6 countries, each with 2 properties= 12
for(i=0;i<size;i++)
    {
        Vinv[i]=new double[size];

```

```

        for(j=0;j<size;j++)
Vinv[i][j]=Vinv_cmat.get_element(i,j).Re();
    }
    double **V=inverse_matrix(Vinv,size);

    delete [] lambda_complex;
    doubledelete(Vinv,size);

    return V;
}

double lambda1,lambda2;

// minusloglik=-log(likelihood):
// inputs parameters (reparametrized to go from -infty to infty):
double minusloglik(double *pars)
{
    // Read the parameters from the array (make sure to parametrized them
    // back to their original values):

    double s1=exp(pars[0]);
    double s2=exp(pars[1]);
    double aI=-5.0+10.0*(1.0/(1.0+exp(-pars[2])));
    double aG=-5.0+10.0*(1.0/(1.0+exp(-pars[3])));
    double alphaI=(1.0/(1.0+exp(-pars[4])));
    double alphaG=(1.0/(1.0+exp(-pars[5])));
    double sigmaI=exp(pars[6]);
    double sigmaG=exp(pars[7]);

    // check if the parameters have legal values:
    int e=0;
    // check if the exponential of the parameters are within the
    // limit (for the pull, diffusion and interaction parameters)
    for (e=0; e<8; e++)

```

```

    {
        if(!(pars[e] > -1e+200 && pars[e]<1e+200))
return 1e+200;
    }

// index variables:
int j,k,l, i;

// Make pull matrix, expectation vector and diffusion matrix
// specific to our 6-country-2-properties-and-2-underlying processes case

int size=12; //6 countries times 2 properties + 2 underlying

int minctr=meas[0].ctrnr;
for(i=1;i<meas_len;i++)
    if(meas[i].ctrnr<minctr)
        minctr=meas[i].ctrnr;

// Make the pull matrix:
double **A=new double*[size];
for(i=0;i<size;i++)
{
    A[i]=new double[size];

    for(j=0;j<size;j++)
A[i][j]=0.0;
}
// fill the pull matrix:
for(i=0;i<6;i++)
{
    A[2*i][2*i]= -alphaI;
    A[2*i][2*i+1]= alphaI*aG;

```

```

        A[2*i+1][2*i+1]=-alphaG;
        A[2*i+1][2*i]=alphaG*aI;
    }

    // Make the expectancy vector:
    double *m0=new double[size];
    for (i=0; i<size; i++)
        m0[i]=0;

    // Make the Sigma2 matrix:
    double **Sigma2=new double*[size];
    for(i=0;i<size;i++)
    {
        Sigma2[i]=new double[size];

        for(j=0;j<size;j++)
            Sigma2[i][j]=0.0;
    }
    // fill the Sigma2 matrix:
    for(i=0;i<6;i++)
    {
        Sigma2[2*i][2*i]=sigmaI*sigmaI;
        Sigma2[2*i+1][2*i+1]=sigmaG*sigmaG;
    }

    // Find eigenvalues and eigenvectors:
    double *lambda;
    double **V=get_eigenvector_matrix(A,size,&lambda);
    lambda1=lambda[0];
    lambda2=lambda[1];

    if(lambda1>0.0 || lambda2>0.0)
        return 1e+200;

```

```

if(V==NULL) // eigenvectors or -values were complex?
    return 1e+200;
double **Vinv=inverse_matrix(V,size);

// VSigma2=V*sigma^2
double **VSigma2=new double*[size];
for(i=0;i<size;i++)
{
    VSigma2[i]=new double[size];
    for(j=0;j<size;j++)
{
    VSigma2[i][j]=0.0;
    for(k=0;k<size;k++)
        VSigma2[i][j]+=V[i][k]*Sigma2[k][j];
    }
}

// Omega=V*sigma^2*V'
double **Omega=new double*[size];
for(i=0;i<size;i++)
{
    Omega[i]=new double[size];
    for(j=0;j<size;j++)
{
    Omega[i][j]=0.0;
    for(k=0;k<size;k++)
        Omega[i][j]+=VSigma2[i][k]*V[j][k];
    }
}

// make the covariance matrix:
double **covar=new double*[2*meas_len];
for(i=0;i<(2*meas_len);i++)
    covar[i]=new double[2*meas_len];

```

```

// make the expectancy vector:
double *mean=new double[2*meas_len];

// make the observation vector:
double *y=new double[2*meas_len];

// Traverse the measurements and fill out the expectancy and
// observation vector:
for(i=0;i<meas_len;i++)
{
    // observations: first all inequalities, thereafter all generousities,
    // in an array of length 2*meas_length
    y[i]=meas[i].ineq;
    y[i+meas_len]=meas[i].gen;

    int index=meas[i].ctrnr-minctr;

    // expectations:
    mean[i]=0.0;
    /*
    for(j=0;j<size;j++)
for(k=0;k<size;k++)
    mean[i] += Vinv[2*index][j]*(-1.0/lambda[j])*V[j][k]*m0[k];
    */

    // expectations:
    mean[i+meas_len]=0.0;
    /*
    for(j=0;j<size;j++)
for(k=0;k<size;k++)
    mean[i+meas_len] +=
Vinv[2*index+1][j]*(-1.0/lambda[j])*V[j][k]*m0[k];
    */
}

```

```

//print out observations and expectations:
//for (i=0; i<meas_len; i++)
// printf("%lf\t %lf\t\t %lf\t %lf\n", y[i], mean[i], y[i+meas_len], mean[i+meas_len]

// fill out the covariance matrix:
for(i=0;i<meas_len;i++)
    for(j=0;j<meas_len;j++)
        {
// initialize:
covar[i][j]=0.0;
covar[i+meas_len][j]=0.0;
covar[i][j+meas_len]=0.0;
covar[i+meas_len][j+meas_len]=0.0;

// If observation i and j are the same, add measurement noise:
if(i==j)
    {
        covar[i][j]+=s1*s1;
        covar[i+meas_len][j+meas_len]+=s2*s2;
    }

// find the absolute time difference between
// measurement i and j:
double t=-meas[i].year, v=-meas[j].year;
double tdiff=ABSVAL((t-v));

// we are looking at the top layer, at process number
// "meas[i].site" and "meas[i].site"
int index1=meas[i].ctrnr-minctr;
int index2=meas[j].ctrnr-minctr;

if(index1==index2)
    {

```

```

// covar = (V^-1) Omega/(-lambda1-lambda2) exp(lambda*(t-v)) (V^-1)'
for(k=0;k<size;k++)
    for(l=0;l<size;l++)
if (ABSVAL((Omega[k] [l]))>0.0001)
{
    covar[i] [j] +=
        Vinv[2*index1] [k]*(-Omega[k] [l]/(lambda[k]+lambda[l]))*
        exp(lambda[l]*tdiff)*Vinv[2*index2] [l];

    covar[i+meas_len] [j] +=
        Vinv[2*index1+1] [k]*(-Omega[k] [l]/(lambda[k]+lambda[l]))*
        exp(lambda[l]*tdiff)*Vinv[2*index2] [l];

    covar[i] [j+meas_len] +=
        Vinv[2*index1] [k]*(-Omega[k] [l]/(lambda[k]+lambda[l]))*
        exp(lambda[l]*tdiff)*Vinv[2*index2+1] [l];

    covar[i+meas_len] [j+meas_len] +=
        Vinv[2*index1+1] [k]*(-Omega[k] [l]/(lambda[k]+lambda[l]))*
        exp(lambda[l]*tdiff)*Vinv[2*index2+1] [l];

}
}
}

// Calculate the negative log-likelihood:
double ret = - multinormal_pdf(y, mean, covar, 2*meas_len, b_true, b_true);

// cleanup:
doubledelete(covar,2*meas_len);
delete [] y;
delete [] mean;
doubledelete(A,size);
doubledelete(V,size);

```



```

doubledelete(Vinv,size);
doubledelete(Sigma2,size);
doubledelete(VSigma2,size);
doubledelete(Omega,size);
delete [] lambda;
delete [] m0;

//return the negative log-likelihood
return ret;
}

void show_parameters(double *pars, double l1, double l2)
{
    double ml=minusloglik(pars);
    double s1=exp(pars[0]);
    double s2=exp(pars[1]);
    double aI=-5.0+10.0*(1.0/(1.0+exp(-pars[2])));
    double aG=-5.0+10.0*(1.0/(1.0+exp(-pars[3])));
    double alphaI=(1.0/(1.0+exp(-pars[4])));
    double alphaG=(1.0/(1.0+exp(-pars[5])));
    double sigmaI=exp(pars[6]);
    double sigmaG=exp(pars[7]);

    // show the estimated parameters on the screen:
    printf("-loglik:  %f\n",ml);
    printf(" s1=%f s2=%f aI=%f aG=%f alphaI=%f alphaG=%f \n "
"sigmaI=%f sigmaG=%f -loglik=%f\n",
s1, s2, aI, aG, alphaI, alphaG, sigmaI,
sigmaG, ml);
    printf("lambda1=%f lambda2=%f\n", lambda1, lambda2);
    printf("dt1=%f dt2=%f\n", -1.0/lambda1, -1.0/lambda2);
}

int main(int argc, char **argv)

```

```

{
    // check the number of user inputs:
    if(argc!=3)
        usage();

    // fetch the measurements:
    meas=get_measurements(argv[1],&meas_len);

    // initialize the random seed:
    randify();

    // fetch the user-specified number of hill-climbs:
    int numstart=atoi(argv[2]);

    // index:
    int i,j;

    // best log-likelihood:
    double min_minusloglik;

    // make vectors for current and best parameter values
    // (re-parametrized to go from minus infty to infty):
    double *pars=new double[8], *best_pars=new double[8];
    double best_lambda1, best_lambda2;

    // traverse the number of wanted hill-climbs:
    for(i=0;i<numstart;i++)
    {
        double before_minusloglik;

        // initialize the starting parameter values randomly:
        do
        {
            for(j=0;j<8;j++)
                pars[j]=-4.0+8*drand48();

```

```

before_minusloglik=minusloglik(pars);

//printf("Loop:  %f\n",before_minusloglik);
} while(!(before_minusloglik<1e+199));

printf("Før:\n");
show_parameters(pars,lambda1,lambda2);

// do the optimization:
int maxiter=1000;
double *pars2=gsl_optimization_cover(minusloglik,
8 /* number of parameters */,
pars /* starting values */,
0.0001 /* precision */,
maxiter /* Maximum number of
iterations. The number
of iterations
neccessary will be
stored here after the
optimization and can
be checked */);
cout << "iterations:" << maxiter << endl;

double after_minusloglik=minusloglik(pars2);
double before_minusloglik2=minusloglik(pars);

printf("Etter:\n");
show_parameters(pars2, lambda1, lambda2);

// Check if this is the best result so far:
if(i==0 || after_minusloglik<min_minusloglik)
{
// If so, store the parameter array and the log-likelihood:
for(j=0;j<8;j++)

```

```

        best_pars[j]=pars2[j];
min_minusloglik=after_minusloglik;
best_lambda1=lambda1;
best_lambda2=lambda2;
}

        delete [] pars2;
    }

printf("\n\nBest:\n");
show_parameters(best_pars, best_lambda1, best_lambda2);

// quit:
return 0;
}

```

## Referanser

- Gosta Esping-Andersen. *Three Worlds of Welfare Capitalism*. Princeton University Press, 1990. ISBN 9780691028576.
- Barth Erling Moene, Karl Ove. The equality multiplier. *unpublished*, 2009.
- Schweder T Henderiks J Reitan, T. Modelling evolution by stochastic differential euqations with latent layers. *unpublished*, 2010.
- A Seierstad. *Stochastic Control in Discrete and Continous Time*. Springer-Verlag, Berlin, 2009. ISBN 9780387766164.
- Berck Sydsaeter, Stroem. *Matematisk formelsamling for okonomer*. Gyldendal, Oslo, 2006a. ISBN 9788205366213.
- Oksendal Sydsaeter. *Lineaer Algebra*. Gyldendal, Oslo, 2006b. ISBN 9788200223696.

Scruggs, Lyle: Comparative Welfare Entitlements Dataset: Generosity Index:  
<http://sp.uconn.edu/~scruggs/cwed/generosity12.xls>

## 9 Etterord

Det har ikke vært en enkel oppgave å jobbe med. Hadde det ikke vært for min biveileder, Trond Reitan, ville det heller ikke gått. Det ville iallefall stoppet før numerikken, og da ville mye av poenget være borte. Optimeringen er gjort basert på en modifisert kode av en optimering han tidligere har gjort. Hjelpen jeg har fått av han har vært et kurs i C separat, og det har vært særdeles interessant. Etter at min veileder, Tore Schweder, nå på slutten har gjennomlest min oppgave, har det medført betydelige endringer. Graden av diskusjon er øket, og mange uklarheter er fjernet. I denne siste diskusjonen har nye problemstillinger dukket opp, og tidligere feilvurderinger har blitt helt klare. Det er en hel del jeg ville gjort annerledes om jeg begynte igjen idag. Jeg ville heller begynne med enkle modeller, for deretter gradvis å bygge inn kompliserende elementer, samtidig som dette ble satt i et økonomisk perspektiv. Men likevel har det vært en særdeles lærerik prosess. Mye har jeg lært om stokastisk analyse, og enda mer har jeg lært om hvordan skrive, og ikke mindre hvordan ikke skrive, oppgave. Mye av denne prosessen har vært preget av usikkerhet. Jeg visste ofte ikke hva jeg ikke visste at jeg måtte vite. Horisonten har dermed vært for kort, for jeg visste ikke helt hva jeg kunne besvare før jeg var helt fortrolig med selve modellrammeverket. Jeg visste heller ikke helt hva jeg kunne få besvart via numerisk metode før jeg visste hvordan dette fungerte. Slik ville jeg ikke gått frem igjen idag. Et visst bilde av sluttproduktet tror jeg er en god forutsetning for å gå i gang med arbeidet. Altså ikke nødvendigvis, eller gjerne nødvendigvis ikke, et bilde over hva konklusjonen kommer til å bli, men av hvilken natur den vil være av.

Jeg ser frem til å gå igang med nye problemstillinger på et senere tidspunkt. Denne oppgaven har i stor grad gjort meg mer klar for nettopp dette, fordi jeg i større grad vet hva noe slikt *kan* gå ut på.

Jeg har hatt to strålende veiledere; Tore Schweder som hovedveileder, og

Trond Reitan som biveileder. Jeg unner enhver å få dem som veiledere på et senere tidspunkt. Diskusjonene med dem, læringen og hjelpen, alt har vært utmerket, og jeg takker.